

Corpus of Linguistically Significant Entities (CLSE)

Version: 1 Nov 2022

Author: Aleksandr Chuklin

The Google Knowledge Graph ([KG](#)) has millions of entries that describe real-world entities like people, places, and things. These entities form the nodes of the graph.

For entities in the knowledge graph, we source lexical annotations from expert linguists. Lexical annotations are language-specific and pertain to broader categories of linguistic properties like Animacy, Case, Classifier, Countability, Definiteness, Gender, and Number. Each language uses different linguistic properties. For example, the concept of animacy is not used in the English language.

lang	name	signature	semantic type
fr	Suisse	number:SINGULAR,gender:FEMININE, starts_w_vowel:0	Country
de	Champions League	number:SINGULAR,gender:FEMININE, article:DEFINITE, locative_prep:PREP_IN, directional_prep:PREP_NACH	Event
ru	Саратовские авиалинии	number:PLURAL,casus:NOMINATIVE, allative:PREP_K, comitative:PREP_S, topical:PREP_O, locative_prep_geo:PREP_V	Corporation

Linguistic annotations for an entity include those that are important to handle in a template-based language generation context. For instance in English, location entities have locative preposition annotations while people entities have gender annotations.¹ In other languages like French, *all* entities are annotated for grammatical gender and entities with an article are marked depending on whether this article stays unchanged or gets merged with a preposition like it would for common nouns.

The traffic in Dallas.
The traffic on 101.
The traffic on the Golden Gate Bridge.

Naomi Osaka is 24 years old.
She's 24 years old.

¹ Note that the gender annotations may be sometimes incomplete or inaccurate due to changed state of the world, an annotator mistake, or a lack of standard linguistic handling for gender non-binary persons in certain languages.

Queen Victoria had 9 children.
 Queen Elizabeth has 4 children.
 Carmelo Anthony has 1 child.

Golden State won against the Clippers yesterday 115-113.
The Warriors are currently 1st in the Western Conference.

Figure 1. Linguistic annotations are organized around NLG applications.

We introduce the concept of a **linguistic signature**, which is a linearized string representation of an entity’s linguistic attributes for a specific language. See Figure 1 for an example of this. The table below summarizes linguistic signatures found in the dataset.

The maximum hypothetical number of distinct linguistic signatures for a language is the cartesian product of all linguistic features and values for that language. However, not all linguistic signatures are naturally occurring or relevant. For example, mass nouns that start with a vowel do not occur in Spanish.

To obtain entities based on linguistic variation, we annotate a large number of entities for each semantic type to create a table of (language, semantic type, mid, name, linguistic_signature). We group rows in the table by (language, semantic type, linguistic_signature) triples, and select up to ten entities to represent each triple. This gives us a collection of up to $10 * |\text{linguistic signatures}|$ distinct entities for a given language and semantic type, balanced by linguistic signature. The complete corpus covers 34 languages.

language	ar	bn	cmn-CN	cmn-TW	cs	da	de	en	es	fr	gu	
# unique entities	899	721	530	529	1238	1286	2922	4076	3181	4312	798	
# ling. attributes	21	6	5	6	20	23	37	32	26	26	6	
language	hi	id	it	ja	jv	kn	ko	ml	mr	nl	no	
# unique entities	950	705	2510	1063	55	849	885	888	924	1049	1237	
# ling. attributes	18	9	41	47	2	6	9	6	6	20	23	
language	pl	pt	ru	su	sv	ta	te	th	tr	ur	vi	yue
# unique entities	1606	2464	3039	29	1309	891	885	724	1262	883	612	551
# ling. attributes	30	25	31	2	19	6	6	9	10	6	8	19

Table 1. List of languages in CLSE and # of unique annotated entities for each language.

There are other linguistically significant phenomena that arise from non-entities like verbs and numbers. We leave the extension of this corpus to other grammar-affecting categories to future work.

Please, cite the following work when using the dataset:

```
@inproceedings{clse2022,
  title={CLSE: Corpus of Linguistically Significant Entities},
  author={Chuklin, Aleksandr and Zhao, Justin and Kale, Mihir},
  booktitle={Proceedings of the 2nd Workshop on Natural Language Generation, Evaluation, and Metrics (GEM 2022) at EMNLP 2022},
  year={2022}
}
```

Linguistic attributes:

Name	short description
CommonNoun.countability:BIPARTITE	nouns that have two parts (e.g., scissors, pants)
CommonNoun.countability:COUNT	count noun
CommonNoun.countability:MASS	mass noun
CommonNoun.locative_preposition:PRE_P_A	locative case is formed with a preposition "a" or a diacriticized variants thereof
CommonNoun.mass_count:COUNT	count noun
CommonNoun.mass_count:MASS	mass noun
Determiner.form:DETERMINER_NONE	where there should not be a determiner added, generally because it's a proper noun that doesn't require one either because it's something like a person or place name or because there is already a determiner included in the name (e.g., The Weeknd)
Determiner.form:FORM_THE	where there should be a determiner added, and it'd be "the"
Determiner.mass_count:COUNT	this is treated as count noun for the purposes of determiner handling
InflectedNounForm.casus:ANY_CASUS	Casus is not marked. Likely the same surface form is used for different grammatical cases.
InflectedNounForm.casus:GENITIVE	"name" field contains a genitive form instead of nominative. Likely a data mistake (only applies to 3 mids in Danish).
InflectedNounForm.casus:NEUTRAL	Lack of case marking (e.g. sukkon ending) or default form. (Arabic)
InflectedNounForm.casus:NOMINATIVE	Nominative case (usually the default form for a noun in most languages)

InflectedNounForm.definiteness:DEFINITE	Forms which are morphologically or semantically definite, e.g. proper nouns For common nouns this include forms which only have a definite reading, ignoring deixis. E.g. "hunden" in Danish (Norwegian, Danish, and Swedish)
InflectedNounForm.definiteness:INDEFINITE	Forms which only have an indefinite reading (Norwegian and Swedish only)
InflectedNounForm.definiteness:NEUTRAL	forms which typically have an indefinite reading, but can also be used with a definite article (e.g. "en ø" but also "den ø"). Danish-only
InflectedNounForm.number:PLURAL	the noun should be treated as plural
InflectedNounForm.number:SINGULAR	the noun should be treated as singular
InflectedNounForm.state:CONSTRUCT_STATE	The state field is intended for bare nouns (or noun phrase) which can inflect and be either definite (al-baytu), indefinite (baytu-n), or construct (baytu). (Arabic only)
InflectedNounForm.state:DEFINITE_STATE	The state field is intended for bare nouns (or noun phrase) which can inflect and be either definite (al-baytu), indefinite (baytu-n), or construct (baytu). (Arabic only)
InflectedNounForm.state:INDEFINITE_STATE	The state field is intended for bare nouns (or noun phrase) which can inflect and be either definite (al-baytu), indefinite (baytu-n), or construct (baytu). (Arabic only)
NameTags.appellation:AUTONYM	The dimension for ways to name people or groups thereof that are different from their official given name. AUTONYM: The actual name of this person, whose preferred form is a pseudonym; e.g., "Samuel Clemens" for Mark Twain, "Domenico Theotokopoulos" for El Greco.
NameTags.appellation:COLOR	The dimension for ways to name people or groups thereof that are different from their official given name. COLOR: When referring to a name (typically a team) by their color, e.g. 'Les Bleus' for the France team.
NameTags.appellation:MASCOT	The dimension for ways to name people or groups thereof that are different from their official given name. MASCOT: The name of the mascot representing the people, usually as a group or team.
NameTags.appellation:METONYM	The dimension for ways to name people or groups thereof that are different from their official given name. METONYM: The name of the entity given by the location they usually are in. For team names, it could be the name of the city of their home stadium.

NameTags.appellation:PSEUDONYM	The dimension for ways to name people or groups thereof that are different from their official given name. PSEUDONYM: The well-known pseudonym of this entity; e.g., "Vernon Sullivan" for Boris Vian, "Richard Bachman" for Stephen King or 'big Apple' for New York City.
NameTags.completeness:AMBIGUOUS	A short name of the entity or person, that can be used in a conversation once the entity has been mentioned before and it becomes clear what is meant by that name. Examples: 'Clinton' after the full name has been used, or 'Paris' after 'Paris Hilton' or 'Paris, Texas' has been mentioned.
NameTags.completeness:UNAMBIGUOUS	A name of the person or entity, which is sufficient to identify it when used for the first time in a conversation, e.g. "Elton John".
NameTags.formality:COLLOQUIAL	Marks a name that is commonly used for the entity, and can be just as identifiable as the official name, e.g. 'Tschechien' or 'Bill Clinton'.
NameTags.formality:OFFICIAL	The name that is referred to as the official name. It tends usually to be a more complete or longer name, e.g. 'Tschechische Republik' or 'William Clinton'.
NameTags.geo:ENDONYM	The full form that would be used when the user is local to it. Eg: "Paris, Texas" for a user in the United States.
NameTags.geo:EXONYM	This is the full form that would be given over a different country than the user's. E.g., "Paris, France" for a user in the US, "Paris, Texas, US" for a user in France.
NameTags.nutrition:PARTITIVE_NUMBER_PREFERENCE	The number-inflected form of a food entity that is typically used in a partitive context for nutrition answers. This allows the nutrition client to answer with the right form, e.g., "...one cup of watermelon (sg)", vs. "one cup of blueberries (pl)".
NameTags.surface:SCREEN	Surface form marked as more suitable to show on a screen (as opposed to spoken).
NameTags.whenis:NO_DATE_NAME	Name referring to a particular dated event (Italian only)
NominalInflectedForm.article_handling: ADD	Behaves like a common noun and uses an article: "L'Alsace" (Spanish, Portuguese, French)
NominalInflectedForm.article_handling: DEFINITE_ARTICLE	Requires definite article (like common nouns).
NominalInflectedForm.article_handling: EMBEDDED_ARTICLE	Article is embedded in the orthography and should be modified accordingly when used in context. (Italian)
NominalInflectedForm.article_handling: MODIFY_EMBEDDED	The orthography does have a capitalized article which behaves like a regular article: "Le Havre requiert ...", "Le port du Havre contient ...", but *not* "La mairie de La Rochelle ouvre ...". (French)
NominalInflectedForm.article_handling: NONE	Most proper nouns do not use an article in a noun phrase. (Spanish, Portuguese, French)

NominalInflectedForm.article_handling: NO_ARTICLE	Does not use an article (most proper nouns).
NominalInflectedForm.gender:FEMININE	Grammatical gender
NominalInflectedForm.gender:MASCULINE	Grammatical gender
NominalInflectedForm.gender:NEUTER	Grammatical gender
NominalInflectedForm.number:PLURAL	Plural/singular form
NominalInflectedForm.number:SINGULAR	Plural/singular form
NominalInflectedForm.override_article:ARTICLE_GLI	This overrides the default article choice for cases where the purely typographic article decision is not sufficient. Default typographic rules are generally insufficient for semivowels /j/ and /w/, and very special cases like "lo Utah". (Italian only)
NominalInflectedForm.override_article:ARTICLE_IL	This overrides the default article choice for cases where the purely typographic article decision is not sufficient. Default typographic rules are generally insufficient for semivowels /j/ and /w/, and very special cases like "lo Utah". (Italian only)
NominalInflectedForm.override_article:ARTICLE_L	This overrides the default article choice for cases where the purely typographic article decision is not sufficient. Default typographic rules are generally insufficient for semivowels /j/ and /w/, and very special cases like "lo Utah". (Italian only)
NominalInflectedForm.override_article:ARTICLE_LA	This overrides the default article choice for cases where the purely typographic article decision is not sufficient. Default typographic rules are generally insufficient for semivowels /j/ and /w/, and very special cases like "lo Utah". (Italian only)
NominalInflectedForm.override_article:ARTICLE_LO	This overrides the default article choice for cases where the purely typographic article decision is not sufficient. Default typographic rules are generally insufficient for semivowels /j/ and /w/, and very special cases like "lo Utah". (Italian only)

NominalInflectedForm.starts_with_phonetic_a:NO	This overrides the default article/preposition choice for cases where the purely typographic article decision is not sufficient.
NominalInflectedForm.starts_with_phonetic_a:YES	This overrides the default article/preposition choice for cases where the purely typographic article decision is not sufficient.
NominalInflectedForm.starts_with_phonetic_e:NO	This overrides the default article/preposition choice for cases where the purely typographic article decision is not sufficient.
NominalInflectedForm.starts_with_phonetic_e:YES	This overrides the default article/preposition choice for cases where the purely typographic article decision is not sufficient.
NominalInflectedForm.starts_with_phonetic_i:NO	This overrides the default article/preposition choice for cases where the purely typographic article decision is not sufficient.
NominalInflectedForm.starts_with_phonetic_i:YES	This overrides the default article/preposition choice for cases where the purely typographic article decision is not sufficient.
NominalInflectedForm.starts_with_phonetic_o:NO	This overrides the default article/preposition choice for cases where the purely typographic article decision is not sufficient.
NominalInflectedForm.starts_with_phonetic_o:YES	This overrides the default article/preposition choice for cases where the purely typographic article decision is not sufficient.
NominalInflectedForm.starts_with_stressed_a:0	This overrides the default article/preposition choice for cases where the purely typographic article decision is not sufficient.
NominalInflectedForm.starts_with_stressed_a:1	This overrides the default article/preposition choice for cases where the purely typographic article decision is not sufficient.
NominalInflectedForm.starts_with_vowel:0	This overrides the default article/preposition choice for cases where the purely typographic article decision is not sufficient.
NominalInflectedForm.starts_with_vowel:1	This overrides the default article/preposition choice for cases where the purely typographic article decision is not sufficient.
Noun.InflectedForm.casus:DIRECT	Direct case (Hindi-only)
Noun.InflectedForm.casus:NOMINATIVE	Nominative case (usually the default form for a noun in most languages)
Noun.InflectedForm.declension:ANY_DECLENSION	Declension not specified / the word does not inflect (German)
Noun.InflectedForm.declension:STRONG	Strong declension. Only set for multi-word names with (e.g., with an adjective) (German)
Noun.InflectedForm.gender:MASCULINE	Grammatical gender

Noun.InflectedForm.number:PLURAL	Plural/singular form
Noun.InflectedForm.number:SINGULAR	Plural/singular form
Noun.allative:PREP_K	Whether preposition "к" or "ко" is used (Russian)
Noun.allative:PREP_KO	Whether preposition "к" or "ко" is used (Russian)
Noun.animacy:ANIMATE	Grammatical animacy of the entity
Noun.animacy:INANIMATE	Grammatical animacy of the entity
Noun.animacy:IRRATIONAL	Grammatical animacy of the entity
Noun.animacy:RATIONAL	Grammatical animacy of the entity
Noun.article_handling:ADD_CAPITALIZED_DEFINITE_ARTICLE	Adding a definite article that is capitalized because the definite article is considered part of the name
Noun.article_handling:DEFINITE_ARTICLE	Noun requires a definite article (e.g., the Eiffel Tower)
Noun.article_handling:INDEFINITE_ARTICLE	Usually used for common count nouns
Noun.article_handling:NO_ARTICLE	Noun does not require the addition of an article (e.g., Yankee Stadium)
Noun.comitative:PREP_S	Whether preposition "с" or "со" is used (Russian)
Noun.comitative:PREP_SO	Whether preposition "с" or "со" is used (Russian)
Noun.countability:COUNT	count noun
Noun.countability:MASS	mass noun
Noun.dative_preposition:PREP_K	Which preposition should be used in Dative (Czech)
Noun.dative_preposition:PREP_KE	Which preposition should be used in Dative (Czech)
Noun.definiteness:DEFINITE	Definite noun (Arabic)
Noun.directional_preposition:PREP_AUF	Which directional preposition should be used (German)
Noun.directional_preposition:PREP_IN	Which directional preposition should be used (German)
Noun.directional_preposition:PREP_NACH	Which directional preposition should be used (German)
Noun.directional_preposition:PREP_ZU	Which directional preposition should be used (German)
Noun.gender:COMMON	Grammatical gender
Noun.gender:FEMININE	Grammatical gender
Noun.gender:IRRELEVANT_PLURALTANTUM	A noun which is plural in form but singular in meaning
Noun.gender:MASCULINE	Grammatical gender
Noun.gender:MASCULINE_ANIMATE	Grammatical gender
Noun.gender:MASCULINE_INANIMATE	Grammatical gender
Noun.gender:MASCULINE_PERSON	Grammatical gender
Noun.gender:NATURAL_GENDER	Grammatical gender
Noun.gender:NEUTER	Grammatical gender

Noun.genitive_preposition:PREP_Z	Which preposition should be used in Genitive (Czech)
Noun.genitive_preposition:PREP_ZE	Which preposition should be used in Genitive (Czech)
Noun.honorificity:INHERENTLY_HONORIFIC	Honorificity (Hindi)
Noun.honorificity:NOT_INHERENTLY_HONORIFIC	Honorificity (Hindi)
Noun.instrumental_preposition:PREP_S	Which preposition should be used in Instrumental case (Czech)
Noun.instrumental_preposition:PREP_SE	Which preposition should be used in Instrumental case (Czech)
Noun.locative_preposition:NO_PREP	Indicates that no preposition should be used in the locative form (German)
Noun.locative_preposition:NO_PREPOSITION	Indicates that no preposition should be used in the locative form (Norwegian, Danish)
Noun.locative_preposition:PREP_AN	Which locative preposition should be used
Noun.locative_preposition:PREP_AUF	Which locative preposition should be used
Noun.locative_preposition:PREP_BEI	Which locative preposition should be used
Noun.locative_preposition:PREP_BIJ	Which locative preposition should be used
Noun.locative_preposition:PREP_I	Which locative preposition should be used
Noun.locative_preposition:PREP_IN	Which locative preposition should be used
Noun.locative_preposition:PREP_NA	Which locative preposition should be used
Noun.locative_preposition:PREP_OP	Which locative preposition should be used
Noun.locative_preposition:PREP_PA	Which locative preposition should be used
Noun.locative_preposition:PREP_V	Which locative preposition should be used
Noun.locative_preposition:PREP_VE	Which locative preposition should be used
Noun.locative_preposition:PREP_VED	Which locative preposition should be used
Noun.locative_preposition:PREP_VID	Which locative preposition should be used
Noun.locative_preposition_geographical:PREP_NA	Which locative preposition should be used (Russian)
Noun.locative_preposition_geographical:PREP_V	Which locative preposition should be used (Russian)
Noun.locative_preposition_geographical:PREP_VO	Which locative preposition should be used (Russian)
Noun.locative_preposition_political:PREP_NA	Which locative preposition should be used (Russian)
Noun.locative_preposition_political:PREP_V	Which locative preposition should be used (Russian)
Noun.locative_preposition_political:PREP_VO	Which locative preposition should be used (Russian)
Noun.mass_count:COUNT	count noun
Noun.mass_count:MASS	mass noun

Noun.noun_type:NOUN_COMMON	Common noun (Hindi)
Noun.noun_type:NOUN_PROPER	Proper noun (Hindi)
Noun.number:PLURAL	Plural/singular form
Noun.number:SINGULAR	Plural/singular form
Noun.preposition_for_in_a_place:PREP_NAD_INSTR	Locative in/to preposition (Polish)
Noun.preposition_for_in_a_place:PREP_NA_LOC	Locative in/to preposition (Polish)
Noun.preposition_for_in_a_place:PREP_POD_INSTR	Locative in/to preposition (Polish)
Noun.preposition_for_in_a_place:PREP_PRZY_LOC	Locative in/to preposition (Polish)
Noun.preposition_for_in_a_place:PREP_W_LOC	Locative in/to preposition (Polish)
Noun.preposition_for_to_a_place:PREP_DO_GEN	Locative in/to preposition (Polish)
Noun.preposition_for_to_a_place:PREP_NAD_ACC	Locative in/to preposition (Polish)
Noun.preposition_for_to_a_place:PREP_NA_ACC	Locative in/to preposition (Polish)
Noun.preposition_for_to_a_place:PREP_POD_ACC	Locative in/to preposition (Polish)
Noun.preposition_for_to_a_place:PREP_W_ACC	Locative in/to preposition (Polish)
Noun.topical:PREP_O	Topical preposition "o"/"o6" (Russian)
Noun.topical:PREP_OB	Topical preposition "o"/"o6" (Russian)
ProperNoun.locative_article_handling:DEFINITE_ARTICLE	There are toponyms for which even though the Np would use an article, the locative form is fossilized into NOT using it (Italian)
ProperNoun.locative_article_handling:NO_ARTICLE	There are toponyms for which even though the Np would use an article, the locative form is fossilized into NOT using it (Italian)
ProperNoun.locative_preposition:PREP_A	Which locative preposition should be used
ProperNoun.locative_preposition:PREP_AT	Which locative preposition should be used
ProperNoun.locative_preposition:PREP_DANS	Which locative preposition should be used
ProperNoun.locative_preposition:PREP_EN	Which locative preposition should be used
ProperNoun.locative_preposition:PREP_IN	Which locative preposition should be used

ProperNoun.locative_preposition:PREP_ON	Which locative preposition should be used
ProperNoun.locative_preposition:PREP_SU	Which locative preposition should be used
ProperNoun.locative_preposition:PREP_SUR	Which locative preposition should be used
SurfaceForm.casus:NOMINATIVE	Nominative case (Polish)
SurfaceForm.gender:MASCULINE_PERSON	Grammatical gender
SurfaceForm.number:PLURAL	Plural/singular form
SurfaceForm.number:SINGULAR	Plural/singular form
Variant.classifier:CLASSIFIER_ACCOUNT	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_AIRCRAFT	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_ANIMAL	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_ANIMAL_BIG	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_ANIMAL_FLYING	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_ANIMAL_SMALL	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_BOOK	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_BUSINESS	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_CANNON	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_CAPSULE	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_CAR	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_CHARACTER	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_CLOTHING_SUIT	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_COMPANY	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_COUNT	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_CUP	Classifier used with this entity in Japanese (ja) or Cantonese (yue)

Variant.classifier:CLASSIFIER_DISTRICT	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_GENERATION	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_GRAVE	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_GUN	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_HOUSE	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_LANDPIECE	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_MACHINE	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_MOVIE	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_NEWSPAPER	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_PARAGRAPH	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_PERSON	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_PERSON_GROUP	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_RAILCAR	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_ROOM	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_SCHOOL	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_SENTEENCE	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_SHIP	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_SIMULACRUM	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_SPORTS_TEAM	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_STOCKS	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_STORY	Classifier used with this entity in Japanese (ja) or Cantonese (yue)

Variant.classifier:CLASSIFIER_SWORD	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_THING	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_THING_FLAT	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_THING_LONG	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_THING_SMALL_ROUND	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_TOOL	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_WALL	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_WARSHIP	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
Variant.classifier:CLASSIFIER_WORD	Classifier used with this entity in Japanese (ja) or Cantonese (yue)
WordStemAnnotation.compound_stem: 0	the word does not end with a compound marker (Turkish)
WordStemAnnotation.compound_stem: 1	the word ends with a compound marker (Turkish)
aliveness:ALIVENESS_DEAD	Annotated aliveness (may not be up-to-date)
aliveness:ALIVENESS_LIVING	Annotated aliveness (may not be up-to-date)
event_type:EVENT_CULTURAL	Semantic event type to ease the verb and tense selection (Portuguese)
event_type:EVENT_HOLIDAY_FIXED	Semantic event type to ease the verb and tense selection (Portuguese)
event_type:EVENT_HOLIDAY_VARIABLE	Semantic event type to ease the verb and tense selection (Portuguese)
event_type:EVENT_ORGANIZED	Semantic event type to ease the verb and tense selection (Portuguese)
event_type:EVENT_ORGANIZED_AND_RECURRENT	Semantic event type to ease the verb and tense selection (Portuguese)
event_type:EVENT_POLITICAL	Semantic event type to ease the verb and tense selection (Portuguese)
event_type:EVENT_TELEVISION	Semantic event type to ease the verb and tense selection (Portuguese)
event_verb:AVVENIRE	Semantic event type to ease the tense selection (Italian)
event_verb:ESSERE	Semantic event type to ease the tense selection (Italian)
event_verb:TENERSI	Semantic event type to ease the tense selection (Italian)
gender:GENDER_FEMALE	Grammatical gender

gender:GENDER_MALE	Grammatical gender
gender:GENDER_NONE	Grammatical gender. GENDER_NONE can be used for non-person entities such as toponyms. Such entities can be safely pronominalized to "it" in English (as opposed to humans with unknown gender).
gender:GENDER_OTHER	Grammatical gender. GENDER_OTHER can be used for people whose gender does not conform to MALE or FEMALE