Laboratoire d’analyse et de technologie du langage Notes techniques 20/x

Spécification de la base de données lexicale

Luka Nerima

Département de linguistique - LATL

Université de Genève

28 mars 2020

**Introduction**

Ce document décrit le schéma de la base de données lexicale et donne une synthèse de l’information linguistique contenue dans cette base de données.

Pour chaque langue, les informations lexicales sont contenues dans trois tables. Ainsi, pour le français, nous avons fr\_word (la table des mots), fr\_lexeme (la table des lexèmes) et fr\_colloc (la table des collocations). La vue fr\_item, jointure sur les tables fr\_word et fr\_lexeme, donne toute les informations d’une entrée lexicale à condition que celle-ci représente sa forme de base. La table universals contient les informations permettant le décodage des champs des tables de la base de données lexicale.

**Schéma de la base**

**fr\_word** key not null varchar2(100)

word\_index not null number(10)

lexeme\_index not null number(10)

cat number(2)

numbers number(1)

gender number(1)

person number(3)

case number(2)

feat number(10)

flags number(10)

freq number(9)

penalty number(9)

phono\_str varchar2(32)

cons\_liaison varchar2(4)

created date

created\_by varchar2(10)

modified date

modified\_by varchar2(10)

**fr\_lexeme**

lexeme\_index not null number(10)

word\_index not null number(10)

next\_lexeme number(10)

idiom\_index number(10)

cat number(2)

inflexion number(10)

type number(2)

subtype number(10)

feat number(10)

sel number(10)

feat3 number(10)

freq number(9)

penalty number(9)

example varchar2(100)

nb\_arg number(1)

a1\_gf number(2)

a1\_theta number(3)

a1\_case number(10)

a1\_type number(2)

a1\_sel\_feat number(10)

a2\_gf number(2)

a2\_theta number(3)

a2\_case number(10)

a2\_type number(2)

a2\_sel\_feat number(10)

a3\_gf number(2)

a3\_theta number(3)

a3\_case number(10)

a3\_type number(2)

a3\_sel\_feat number(10)

a4\_gf number(2)

a4\_theta number(3)

a4\_case number(10)

a4\_type number(2)

a4\_sel\_feat number(10)

created date

created\_by varchar2(10)

modified date

modified\_by varchar2(10)

**fr\_colloc**

colloc\_index not null number(10)

item\_type1 number(1)

lexeme\_index1 number(10)

colloc\_index1 number(10)

item\_type2 number(1)

lexeme\_index2 number(10)

colloc\_index2 number(10)

colloc\_type number(10)

colloc\_feat number(10)

prep number(10)

lexical\_function number(10)

str varchar2(100)

freq number(2)

created date

created\_by varchar2(10)

modified date

modified\_by varchar2(10)

**universals**

code not null varchar2(6)

ord\_value number(4)

type\_id varchar2(20)

lang varchar2(40)

literal varchar2(26)

displ\_literal varchar2(100)

examples varchar2(80)

**La table universals**

Elle est utilisée pour toutes les langues. Elle sert au décodage des champs numériques des tables de la base de données des lexiques, en donnant la correspondance entre les valeurs numériques de trait (ord\_value) et le trait lexical en clair, sous forme de chaîne de caractère (literal), ainsi que la chaîne de caractère à afficher dans l’interface utilisateur du lexique (displ\_literal). Type\_id contient le type du trait, p.ex. *Category*, *Mood*, *SyntacticFeature*, etc. Le champ *lang* contient la liste des langues pour lequel un trait s’applique, p.ex. *fr, en, de* signifie que le trait s’applique au français, anglais et allemand. Si le champ est vide, le trait s’applique à toutes les langues de la base de données lexicale. Le champ *code* est la clé (artificielle) de la table *universals*.

**Vues**

fr\_item est la jointure de fr\_lexeme avec fr\_word, word\_index étant la charnière. Elle sert essentiellement à afficher un lexème avec sa chaîne orthographique.

**fr\_item**

key varchar2(100)

cat number(2)

numbers number(1)

gender number(1)

person number(3)

case number(3)

word\_feat number(10)

flags number(10)

word\_freq number(2)

word\_penalty number(2)

phono\_str varchar2(20)

cons\_liaison varchar2(4)

lexeme\_index not null number(10)

word\_index not null number(10)

next\_lexeme number(10)

idiom\_index number(10)

inflexion number(10)

type number(2)

subtype number(10)

lexeme\_feat number(10)

feat2 number(10)

feat3 number(10)

lexeme\_freq number(2)

lexeme\_penalty number(2)

lexeme\_example varchar2(100)

nb\_arg number(1)

a1\_gf number(2)

a1\_theta number(3)

a1\_case number(10)

a1\_type number(2)

a1\_sel\_feat number(10)

a2\_gf number(2)

a2\_theta number(3)

a2\_case number(10)

a2\_type number(2)

a2\_sel\_feat number(10)

a3\_gf number(2)

a3\_theta number(3)

a3\_case number(10)

a3\_type number(2)

a3\_sel\_feat number(10)

a4\_gf number(2)

a4\_theta number(3)

a4\_case number(10)

a4\_type number(2)

a4\_sel\_feat number(10)

wcreated date

wcreated\_by varchar2(10)

wmodified date

wmodified\_by varchar2(10)

wcreated date

wcreated\_by varchar2(10)

wmodified date

wmodified\_by varchar2(10)

**Informations lexicales pour toutes les langues**

Le champ KEY (table des mots) contient la graphie de l’entrée lexicale (représentation orthographique du mot), le champ PHONO\_STR (table des mots) la représentation phonétique et le champ CAT (table des mots et table des lexèmes) la catégorie lexicale selon les valeurs: 0=N, 1=V, 2=Adj, 3=Prep, 4=Det, 5=Adv, 6=Conj, 7=Interj, 8=T, 9=C, 10=F, 11=Clitic (français), 12=Ponc, 13=Poss (marque du possessif en anglais ) et 13=other (pour les autres langues), 14=CompoundChunk (non-mot: marque une partie initiale d'un mot composé).

Les mots qui sont partie initiale d’un mot composé sont marqué du flag 16 dans le champ FLAGS. Le champ FLAGS contient d’autres informations, comme p.ex. *FormeElidée / FormeNonElidée* pour les mots français.

Les lexemes qui sont impliqués dans des collocations portent le flag 31 dans le champ FEAT3 (table des lexèmes).

Les collocations qui sont elles-mêmes impliquées dans d’autres collocations portent le flag 31 dans le champ COLLOC\_FEAT (table des collocations)

**Français**

Le champ CONS\_LIAISON (table des mots) contient la consonne latente de liaison (s’il y a lieu).

Le champ INFLEXION (table des lexèmes) donne la classe d’inflexion (p.e. classe de conjugaison pour un verbe).

**N**

fr\_word :

numbers {}: singular..plural, Number

gender {}: masculine..feminine, Gender

person {}: 1..6, Person

flags {}:

pronoms: formeElidee..formeNonElidee, Flags

fr\_lexeme :

inflexion T: N\_1..N\_9, NounInflection

type T: (type de nom) pronoun..quantifier, NounType

subtype T:

pronoms: (type de pronom) whPronoun..indefinite, PronounType

quantificateurs: chacun..tous, QuantifierType

feat {}:

pronoms: relativePronoun..dontPronom, PronounFeature

pronoms personnels: weak..mixed, PersonalPronounType T

quantificateurs: (non utilisé) distributiveQ..nonDistributiveQ, QuantifierFeature

sel {}:

pronoms: nominative..distributive , Case -- devrait aller dans fr\_word.case

feat3 {}:

noms communs et noms propres: (traits sémantiques) human..pointCardinal, NounFeature

nb\_arg: (nombre d’arguments) 0..1

a1\_gf T: (fonction grammaticale) subject..Sobject, GrammaticalFunction

a1\_theta T: (rôle thématique) agent..undefined, ThematicFunction

a1\_type T: (type de préposition) timeP..otherP, PrepositionType

a1\_sel\_feat {}:

(traits sélectionnels PP) a..avec, BasicPreposition

(traits sélectionnels CP, ie a1\_gf = SObject) tensedS..subjonctiveS, SFeature

**Clitic**

fr\_word :

numbers {}: singular..plural, Number

gender {}: masculine..neuter, Gender

person {}: 1..6, Person

flags {}: formeElidee..formeNonElidee, Flags

fr\_lexeme :

type T: (type de clitique) enCl..seCl, CliticType

subtype T: (trait de clitique) clMoi..subjInter, CliticFeature

sel {}: nominative..distributive , Case

**V**

fr\_word :

numbers {}: singular..plural, Number

gender {}: masculine..feminine, Gender (pertinent pour les participes passés)

person {}: 1..6, Number

feat {}: (temps et mode) infinitive..participePresentPasse, Tense / indicative..conditional, Mood

fr\_lexeme :

inflexion T: V\_1..V\_4, VerbInflection

type T: (type de verbe) auxiliary..causative, VerbType

subtype : pas utilisé en français

feat {}:

ordinary / causative : (traits syntaxiques / sémantiques),

ergativeV..impersonalV, VerbSynFeature / etatV..genericV, VerbSemFeature,

modal: pouvoir..vouloir, ModalType T

auxiliary: etre..avoir, AuxiliaryType

feat3 {}: NOT YET IMPLEMENTED !

ordinary: (traits sémantiques) etatV..genericV, VerbSemFeature + clivage pour "être"

impersonnel

causative: verbeCausatif + (traits sémantiques) etatV..genericV, VerbSemFeature

auxiliary: 0 + auxiliaireAvoir..auxiliaireEtre

nb\_arg: (nombre d’arguments) 0..3

a1\_gf T: (fonction grammaticale) subject..Sobject, GrammaticalFunction

a1\_theta T: (rôle thématique) agent..undefined, ThematicFunction

a1\_type T: (type de préposition) timeP..otherP, PrepositionType

a1\_sel\_feat {}:

(traits sél. NP, si a1\_gf ∈{subject, directObject, indirectObject}) impersonalNP..bodyPartNP,

NPFeature

(traits sélectionnels PP, si a1\_gf = PrepObject) a..avec, BasicPreposition

(traits sélectionnels CP, si a1\_gf = SObject) tensedS..subjonctiveS, SFeature

(traits sélectionnels FP, si a1\_gf = PredObject) predicativeAP..predicativeDP, PredFeature

idem pour les autres arguments

**Adj**

fr\_word :

numbers {}: singular..plural, Number

gender {}: masculine..feminine, Gender

person {}: 1..6, Person

feat {}: (degré) positive..superlative, Degree

fr\_lexeme :

inflexion : A\_1, A\_12, AdjectiveInflection

feat {}: (traits d’adjectif) toughAdj..perfectParticiple, AdjectiveFeature

nb\_arg: (nombre d’arguments) 0..1

a1\_gf T: (fonction grammaticale) subject..Sobject, GrammaticalFunction

a1\_theta T: (rôle thématique) agent..undefined, ThematicFunction

a1\_type T: (type de préposition) timeP..otherP, PrepositionType

a1\_sel\_feat {}:

(traits sélectionnels PP, ie a1\_gf = PrepObject) a..avec, BasicPreposition

(traits sélectionnels CP, ie a1\_gf = SObject) tensedS..subjonctiveS, SFeature

**Prep**

fr\_word :

flags {}: 0..tresFlag, Flags

fr\_lexeme :

type T: (préposition de base) notABasicP, a..avec, BasicPreposition

feat {}: (type de préposition) timeP..intervalP, PrepositionType

sel {}: intervalP..pronominalP, PrepositionSelection

**Det**

fr\_word :

numbers {}: singular..plural, Number

gender {}: masculine..feminine, Gender

person {}: 1..6, Person

feat {}:

{1} (pour les déterminants contractés du, des, au et aux )

flags {}: (forme élidée/non-élidée) formeElidee..formeNonElidee, Flags + pasUnFlag 21 (41 ds FipsConsts)

fr\_lexeme :

type T: (type de déterminant) definiteDet..comparativeDet, DeterminerType (ord val = ord val + 10)

subtype T:

valeur ordinale pour nombres

personne de référence pour les possessifs : 1..6

(type de déterminant contracté) deDet..auxDet, ContractedDetType

feat {}: ex: ’pas un’ porte le trait polarityItemDet

sel {}: Ncpl..Pde, DeterminerSelection, ex: ’beaucoup’ porte le trait Pde

**Adv**

fr\_word :

flags {}: (forme élidée/non-élidée) formeElidee..formeNonElidee, Flags

fr\_lexeme :

type T: (type d’adverbe) advNegation..advSentential, AdverbType

subtype T: (valeur) vComparison..vFrequency, Value (ord val = ord val + 32)

feat {}: (adverbe complexe/locution adv., inversion du sujet, etc.) advInversion..advCPque, AdverbFeature T

**Conj**

fr\_word :

flags {}: (forme élidée/non-élidée) formeElidee..formeNonElidee, Flags

fr\_lexeme :

type T: (type de conjonction) coordination..subordination, ConjunctionType

subtype T: (valeur de la conjonction) vComparison..vFrequency, Value (ord val = ord val + 32)

feat {}: (traits de conjonction) noInfinitivalConj..enConj, ConjunctionFeature

**Interj**

fr\_lexeme :

type T: (type d’interjection) noCompl..complDpPp, InterjectionType

**Anglais**

Pour l’anglais ce document ne signale que les différences par rapport au français, à savoir FPH (pas de consonne latente en anglais), FLAGS (pas d’élision en anglais), pas de CLITIC en anglais et les différences ci-dessous:

**N**

en\_word :

numbers {}: singular..plural, Number

gender {}: masculine..neuter, Gender

person {}: 1..6, Person

en\_lexeme :

inflexion T: standard..invariable, NounInflection

type T: (type de nom) pronoun..quantifier, NounType

subtype T:

pronoms: (type de pronom) whPronoun..indefinite, PronounType

quantificateurs: everyone..neither, QuantifierType

feat {}:

pronoms: relativePronoun..dontPronom, PronounFeature

pronoms personnels: weak..mixed, PersonalPronounType T

quantificateurs: (non utilisé) distributiveQ..nonDistributiveQ, QuantifierFeature

sel {}:

pronoms: nominative..distributive , Case -- devrait aller dans fr\_word.case

feat3 {}:

noms communs et noms propres: (traits sémantiques) human..pointCardinal, NounFeature

nb\_arg: (nombre d’arguments) 0..1

a1\_gf T: (fonction grammaticale) subject..Sobject, GrammaticalFunction

a1\_theta T: (rôle thématique) agent..undefined, ThematicFunction

a1\_type T: (type de préposition) timeP..otherP, PrepositionType

a1\_sel\_feat {}:

(traits sélectionnels PP) a..avec, BasicPreposition

(traits sélectionnels CP, ie a1\_gf = SObject) tensedS..subjonctiveS, SFeature

**Clitic**

No clitics in English

**V**

en\_word :

numbers {}: singular..plural, Number

gender {}: not used in English

person {}: 1..6, Number

feat {}: (temps) infinitive..conditional, Tense

en\_lexeme :

inflexion T: regular..irregular, VerbInflexion

type T: (type de verbe) auxiliary..causative, VerbType

subtype : (pour les verbes à particule, noParticle si sans particule ) aboutPtl..upPtl, Particle

feat {}:

ordinary / causative : (traits syntaxiques / sémantiques),

ergativeV..impersonalV, VerbSynFeature / etatV..genericV, VerbSemFeature,

modal: can..will, ModalType T

auxiliary: be..let, AuxiliaryType T

feat3 {}: NOT YET IMPLEMENTED !

ordinary: (traits sémantiques) etatV..genericV, VerbSemFeature + clivage pour "être" impersonnel

causative: verbeCausatif + (traits sémantiques) etatV..genericV, VerbSemFeature

auxiliary: 0 + auxiliaireAvoir..auxiliaireEtre

nb\_arg: (nombre d’arguments) 0..3

a1\_gf T: (fonction grammaticale) subject..Sobject, GrammaticalFunction

a1\_theta T: (rôle thématique) agent..undefined, ThematicFunction

a1\_type T: (type de préposition) timeP..otherP, PrepositionType

a1\_sel\_feat {}:

(traits sél. NP, si a1\_gf ∈{subject, directObject, indirectObject}) impersonalNP..bodyPartNP, NPFeature

(traits sélectionnels PP, si a1\_gf = PrepObject) a..avec, BasicPreposition

(traits sélectionnels CP, si a1\_gf = SObject) tensedS..subjonctiveS, SFeature

(traits sélectionnels FP, si a1\_gf = PredObject) predicativeAP..predicativeDP, PredFeature

idem pour les autres arguments

**Adj**

en\_word :

numbers {}: singular..plural, Number

gender {}: not used in English

person {}: 1..6, Person

feat {}: (degré) positive..superlative, Degree

en\_lexeme :

inflexion : A\_1, A\_3, AdjectiveInflection

feat {}: (traits d’adjectif) toughAdj..perfectParticiple, AdjectiveFeature

nb\_arg: (nombre d’arguments) 0..1

a1\_gf T: (fonction grammaticale) subject..Sobject, GrammaticalFunction

a1\_theta T: (rôle thématique) agent..undefined, ThematicFunction

a1\_type T: (type de préposition) timeP..otherP, PrepositionType

a1\_sel\_feat {}:

(traits sélectionnels PP, ie a1\_gf = PrepObject) a..avec, BasicPreposition

(traits sélectionnels CP, ie a1\_gf = SObject) tensedS..subjonctiveS, SFeature

**Prep**

fr\_word :

flags {}: 0..pronunciationFlag, Flags

en\_lexeme :

type T: (préposition de base) notABasicP, about..with, BasicPreposition

subtype T: (si particleP  feat) aboutPtl..upPtl, Particle

feat {}: (type de préposition) timeP..intervalP, PrepositionType

sel {}: intervalP..pronominalP, PrepositionSelection

**Det**

en\_word

numbers {}: singular..plural, Number

gender {}: masculine..neuter, Gender

person {}: 1..6, Person

feat {}: 0 (pas d’articles contractés en anglais)

flags {}: 0..pronunciationFlag, Flags

en\_lexeme

type T: (type de déterminant) definiteDet..comparativeDet, DeterminerType (ord val = ord val + 10)

subtype T:

valeur ordinale pour nombres

personne de référence pour les possessifs : 1..6

(type de déterminant contracté) deDet..auxDet, ContractedDetType

feat {}: ex: ’pas un’ porte le trait polarityItemDet

sel {}: Ncpl..Pde, DeterminerSelection, ex: ’beaucoup’ porte le trait Pde

**Adv**

en\_word :

flags {}: 0..pronunciationFlag, Flags

en\_lexeme :

type T: (type d’adverbe) advNegation..advSentential, AdverbType

subtype T: (valeur) vComparison..vFrequency, Value (ord val = ord val + 32)

feat {}: (adverbe complexe/locution adv., inversion du sujet, etc.) advInversion..advCPque, AdverbFeature T

**Conj**

en\_word :

flags {}: 0..pronunciationFlag, Flags

en\_lexeme :

type T: (type de conjonction) coordination..subordination, ConjunctionType

subtype T: (valeur de la conjonction) vComparison..vFrequency, Value (ord val = ord val + 32)

feat {}: (traits de conjonction) noInfinitivalConj..enConj, ConjunctionFeature

**Interj**

en\_word :

flags {}: 0..pronunciationFlag, Flags

en\_lexeme:

type T: (type d’interjection) noCompl..complDpPp, InterjectionType

**Grec**

Pour le russe ce document ne signale que les différences par rapport au français, à savoir CONS\_LIAISON (pas de consonne latente en russe), FLAGS (pas d’élision en russe) et les différences ci-dessous:

**N**

el\_word :

case {}: (cas ) nominative..genitive, Case

**Clitic**

ras

**V**

el\_word :

feat {}: (temps, mode, aspect et voix) infinitive..futur, Tense / indicative..gerundive, Mood   
 perfectif, imperfectif, Aspect / active, passive, Voice

**Adj**

el\_word :

case {}: (cas ) nominative..genitive, Case

feat {}: pas de degré en grec

**Prep**

ras

**Det**

ras

**Adv**

ras

**Conj**

ras

**Interj**

ras

**Roumain**

Pour le roumain ce document ne signale que les différences par rapport au français, à savoir CONS\_LIAISON (pas de consonne latente en roumain), FLAGS (pas d’élision en roumain) et les différences ci-dessous:

**N**

ro\_word :

case {}: (cas / déclinaison ) nominative..distributive, Case / definiteDecl, indefiniteDecl, DeclensionType

**Clitic**

ras

**V**

ras

**Adj**

ro\_word :

case {}: (cas / déclinaison ) nominative..distributive, Case / definiteDecl, indefiniteDecl, DeclensionType

feat {}: (forme adjectivale) prenominalAdjForm, postnominalAdjForm, AdjectiveForm

**Prep**

ras

**Det**

ras

**Adv**

ras

**Conj**

ras

**Interj**

ras

**Russe**

Pour le russe ce document ne signale que les différences par rapport au français, à savoir CONS\_LIAISON (pas de consonne latente en russe), FLAGS (pas d’élision en russe) et les différences ci-dessous:

**N**

ru\_word :

case {}: (cas / déclinaison ) nominative..instrumental, Case / animate,inamimate, DeclensionType

**Clitic**

ras

**V**

ru\_word :

feat {}: (temps, mode et aspect) infinitive..participePresentPasse, Tense / indicative..conditional, Mood /  
 perfectif, imperfectif, Aspect / short, long, AdjectiveForm (participle only)

**Adj**

ru\_word :

case {}: (cas / déclinaison ) nominative..instrumental, Case / animate,inamimate, DeclensionType

feat {}: (degré / forme ) positive..superlative, Degree / short, long, AdjectiveForm

**Prep**

ras

**Det**

ru\_word :

case {}: (cas / déclinaison ) nominative..instrumental, Case / animate,inamimate, DeclensionType

**Adv**

ras

**Conj**

ras

**Interj**

ras

**Polonais**

Pour le polonais ce document ne signale que les différences par rapport au français, à savoir CONS\_LIAISON (pas de consonne latente en russe), FLAGS (pas d’élision en russe) et les différences ci-dessous:

**N**

pl\_word :

case {}: (cas / déclinaison ) nominative..instrumental, Case / animate,inamimate, DeclensionType

**Clitic**

ras

**V**

pl\_word :

feat {}: (temps, mode et aspect) infinitive..participePresentPasse, Tense / indicative..conditional, Mood /  
 perfectif, imperfectif, Aspect / short, long, AdjectiveForm (participle only)

**Adj**

ru\_word :

case {}: (cas / déclinaison ) nominative..instrumental, Case / animate,inamimate, DeclensionType

feat {}: (degré / forme ) positive..superlative, Degree / short, long, AdjectiveForm

**Prep**

ras

**Det**

pl\_word :

case {}: (cas / déclinaison ) nominative..instrumental, Case / animate,inamimate, DeclensionType

**Adv**

ras

**Conj**

ras

**Interj**

ras

**Allemand**

Le champ INFLEXION (table des lexèmes) donne la classe d’inflexion (p.e. classe de conjugaison pour un verbe).

**N**

de\_word :

numbers {}: singular..plural, Number

gender {}: masculine..neuter, Gender

person {}: 1..6, Person

case {}: nominative..distributive, Case

de\_lexeme :

type T: (type de nom) pronoun..quantifier, NounType

subtype T:

pronoms: whPronoun..indefinite, PronounType T

personnels: weak..strong, PersonalPronounType T

quantificateurs: jedermann..welcher, QuantifierType T

feat {}:

pronoms: relativePronoun..floatingQuantifier, PronounFeature

sel {}:

pronoms: nominative..distributive , Case

feat3 {}:

noms communs et noms propres: (traits sémantiques) human..pointCardinal, NounFeature

**V**

de\_word :

numbers {}: singular..plural, Number

gender {}: masculine..neuter, Gender (pertinent pour les participes passés)

person {}: 1..6, Number

feat {}: (temps/mode) infinitiv..futur2, Tense / indicative..conjunctive, Mood

de\_lexeme :

type T: (type de verbe) auxiliary..causative, VerbType

subtype : (pour verbes à particule) abPtl..zuwiderPtl, Particle

feat {}:

ordinary, causative: ergative..infinitiveWithoutZu / etatV..genericV, VerbSynFeatures / VerbSemFeature

auxiliary: sein..werden, AuxiliaryType

modal: wollen..sollen

**arguments**

nb\_arg: (nombre d’arguments) 0..3

a1\_gf T: (fonction grammaticale) subject..SObject, GrammaticalFunction

a1\_theta T: (rôle thématique) agent..undefined, ThematicFunction

a1\_type T: (type de préposition + 1000 \* cas sélectionné) timeP..causeP, PrepositionType / 1000 \* Case

a1\_sel\_feat {}:

subject, directObject, indirectObject: (sélection NP) nominative..distributive, Case

PrepObject, PrepObject2: (sélection PP) pAn..notABasicP, BasicPreposition

SObject: (sélection CP) tensedS..conjunctiveIndirSpeechS, SFeature

PredObject: (sélection FP) predicativeAP..predicativeDP, PredFeatures

**Adj**

de\_word :

numbers {}: singular..plural, Number

gender {}: masculine..neuter, Gender

person {}: 1..6, Person

case {}: (cas / declension ) nominative..distributive, Case / strong..mixed, DeclensionType

feat {}: (degré) positive..superlative, Degree

de\_lexeme :

feat {}: (traits d’adjectif) toughAdj..perfectParticiple, AdjectiveFeature

nb\_arg: (nombre d’arguments) 0..1

a1\_gf T: (fonction grammaticale) subject..Sobject, GrammaticalFunction

a1\_theta T: (rôle thématique) agent..undefined, ThematicFunction

a1\_type T: (type de préposition + 1000 \* cas sélectionné) timeP..causeP, PrepositionType / 1000 \* Case

a1\_sel\_feat {}:

PrepObject, PrepObject2: (sélection PP) pAn..notABasicP, BasicPreposition

SObject: (sélection S) tensedS..conjunctiveIndirSpeechS, SFeature

**Prep**

de\_word :

flags {}: (forme élidée/non-élidée) formeElidee..formeNonElidee, Flags

de\_lexeme :

type T: (préposition de base) pAn..notABasicP, BasicPreposition

??? subtype: (si la préposition est également une particle) abPtl..zuwiderPtl, Particle

feat {}: (type de préposition / type d’adposition) time..cause, PrepositionType / preposition..contractedPrep,

AdpositionType

feat2 {}: (cas sélectionnés) nominative..distributive, Case

feat3 {}:

circumposition, complexPreposition, doubleComplexPreposition:

(prépositions sélectionnées) pAn..notABasicP, BasicPreposition

**Det**

de\_word :

numbers {}: singular..plural, Number

gender {}: masculine..neuter, Gender

person {}: 1..6, Person

case {}: nominative..distributive, Case

de\_lexeme :

type T: (type de déterminant) definiteDet..comparativeDet, DeterminerType

subtype T:

valeur ordinale pour nombres

personne de référence pour les possessifs : 1..6

feat {}: (type de déclinaison) strong..mixed, DeclensionType

feat2 {}: (sélection) NP..PPvonDefPl, DeterminerSelection

**Adv**

de\_lexeme :

type T: (type d’adverbe) advNegation..advSentential, AdverbType

subtype T: (valeur) vComparison..vUnspecified, Value

**Conj**

de\_lexeme :

type T: (type de conjonction) coordination..subordination, ConjunctionType

subtype T: (valeur de la conjonction) vComparison..vFrequency, Value

feat {}: (traits de conjonction) noInfinitivalConj..enConj, ConjunctionFeature

feat2 {}: multipleConjuncts..CPwhcompl, ConjunctionSelection

**Différences allemand / français**

Idem que français à l’exception de FPH (pas de consonne latente en allemand), FLAGS (pas d’élision en allemand) et des différences ci-dessous:

**N**

de\_word :

case {}: nominative..distributive, Case

de\_lexeme :

feat {}:

pronoms (pas de pronoms clitiques en allemand): whPronoun..indefinite, PronounType T

quantificateurs: jedermann..welcher, QuantifierType T

nb\_arg: 0 (pas d’arguments en allemand)

**V**

de\_lexeme :

subtype : (pour verbes à particule) abPtl..zuwiderPtl, Particle

feat {}:

auxiliary (type du verbe): sein..werden, AuxiliaryType T

modal (type du verbe): wollen..sollen, ModalType T

a1\_type T: (type de préposition + 1000 \* cas sélectionné) timeP..causeP, PrepositionType / 1000 \* Case

a1\_sel\_feat:

subject, directObject, indirectObject: (sélection NP) nominative..distributive, Case

PrepObject, PrepObject2: (sélection PP) pAn..notABasicP, BasicPreposition

**Adj**

de\_word :

case {}: (cas / declension ) nominative..distributive, Case / strong..mixed, DeclensionType

**Prep**

de\_lexeme :

??? subtype: (si particleP  feat) abPtl..zuwiderPtl, Particle

feat {}: (type de préposition / type d’adposition) time..cause, PrepositionType / preposition..contractedPrep,

AdpositionType

feat2 {}: (cas sélectionnés) nominative..distributive, Case

**Det**

de\_word

case {}: nominative..distributive, Case

de\_lexeme

feat {}: (type de déclinaison) strong..mixed, DeclensionType

feat2 {}: (sélection) NP..PPvonDefPl, DeterminerSelection

**Adv**

de\_lexeme:

feat non utilisé

**Conj**

de\_lexeme:

feat2 {}: multipleConjuncts..CPwhcompl, ConjunctionSelection

**Interj**

n’existe pas dans le lexique allemand (pour l’instant)

**Dictionnaires bilingues**

Pour chaque paire de langues, les informations lexicales sont contenues dans une table. Le nom de la table est "bilingual" préfixée par le code ISO (dans l’ordre alphabétique des noms ISO) des deux langues mises en correspondance. Ainsi, pour le lexique bilingue anglais - français, la table s’appelle en\_fr\_ bilingual, pour le français - allemand, de\_fr\_bilingual. Le principe des lexiques bilinques est d’associer deux unités lexicales (généralement deux lexèmes, mais aussi des expressions idiomatiques et des collocations). L’association est complétée par un contexte de sous-langage (p.e. administration, aviation, medcine etc), de correspondance des arguments (p.e. un PP dans une langue devient un NP dans l’autre) et un descripteur pour chaque langue destiné à lever les ambiguïtés dans les cas de lectures multiples. Le schéma des tables est identique pour toutes les langues. La lecture des lexiques bilingues peut s’effectuer dans les deux sens, la première unité lexicale jouant tour à tour le rôle de langage source et de langage source. Tout comme pour les lexiques monolingues, on trouvera dans la table universals les informations permettant la lecture en clair des champs encodés de la table bilingue.

**Schéma des tables bilingues**

**en\_fr\_bilingual**

ITEM\_TYPE1 NUMBER(1)

LEXEME\_INDEX1 NUMBER(9)

IDIOM\_INDEX1 NUMBER(9)

COLLOC\_INDEX1 NUMBER(9)

FEAT\_1 NUMBER(10)

FREQ\_1 NUMBER(3)

DESCRIPTOR\_1 VARCHAR2(40)

ITEM\_TYPE2 NUMBER(1)

LEXEME\_INDEX2 NUMBER(9)

IDIOM\_INDEX2 NUMBER(9)

COLLOC\_INDEX2 NUMBER(9)

FEAT\_2 NUMBER(10)

FREQ\_2 NUMBER(3)

DESCRIPTOR\_2 VARCHAR2(40)

ARG\_MATCH NUMBER(10)

CONTEXT NUMBER(2)

**Information lexicale pour les tables bilingues**

Les champs postfixés par 1 concerne la première langue de la paire, ceux postfixés par 2 la deuxième langue. Dans le préfixe du nom de la table bilingue l’initiale de la langue n°1 précède l’initiale de la langue n°2 et spécifie ainsi quelle est la langue des champs postfixés par 1, respectivement par 2. Ainsi, le nom "e\_f\_bilingual" indique que les champs postfixés par 1 decrivent la partie anglaise de la correspondance alors que les champs postfixés par 2 décrivent la partie française.

Pour la langue n°1

item\_type1: type de l’unité lexicale (1 = lexème, 3 = expression idiomatique, 4 = collocation)

lexeme\_index1,

idiom\_index1,

colloc\_index1: index de l’unité lexicale (un seul des trois est non nul, indiqué par item\_type1)

feat\_1 {}: pas utilisé actuellement

freq\_1: fréquence de la correspondance bilingue dans le sens langue n°1 (source) vers langue n°2 (cible)

descriptor\_1: description textuelle non contenue dans le lexique monolingue destinée à clarifier la lecture de l’unité lexicale

Pour la langue n°2

freq\_2: fréquence de la correspondance bilingue dans le sens langue n°2 (source) vers langue n°1 (cible)

autres champs: idem langue n°1

arg\_match {}: correspondance entre les arguments de la première et de la seconde unité lexicale

context T: (contexte de la traduction au sens sous-langage) Administration..otherContext, ContextType

*Modifications discutées avec Eric et Juri:*

**Modif pour l’allemand**

de\_word:

retour du nombre dans le champs "numbers"

retour de la personne dans le champ "person"

ajout d’un champ "case" pour le cas

degré (degree) dans le champ "feat"

déclinaison (declension) dans le champ "case" (strong=10, weak=11, mixed=12)

N: retour du genre dans le champ "gender"

de\_lexeme

V: subtype particule du verbe (précédement "feat2" et "feat3"). Voir table de\_particle

feat2: traits sélectionnels

**Toutes les langues**

x\_lexeme

Prep: feat2 (cas sélectionnés)

V: ?feat éclaté et placé dans le champ "feat" (synFeat) et "feat2" (semFeat)

**Conversion de l’allemand**

*Ancienne Intermédiaire Nouvelle structure*

n\_infl, v\_infl ... number\_set numbers, gender, person, case,

feat (declension)

**Prep**

prep + basic type: préposition de base ou "notABasicP" idem : BasicPreposition

prep\_type feat{}: PrepositionType idem

prep\_case numbers{}: cas sélectionnés, Case feat2 {}: cas sélectionnés, Case

particle (boolean) featword{}: INCL(10, feat) feat: INCL(10, feat)

nil nil ?subtype T: bePrtl..widerPrtl, ParticleType

**Det**

det\_sel feat {}: DeterminerSelection feat2 {}: DeterminerSelection

decl\_type subtype: declensionType feat {}: DeclensionType

det\_typ type: determinerType idem

per\_ref feat2 {}: personne de réf pour poss., Person subtype: personne de réf pour poss., 1..6

**Conj**

conj\_typ type: type de conjonction, ConjunctionType idem

conj\_sel numbers feat2{}: ConjunctionSelection

conj\_feat\_s feat {}: ConjunctionFeatures idem

conj\_val subtype: valeur de la conjonction, Value idem

**Nom**

genus featlexeme {}: 29(masc),30(fem),31(neuter) gender: Gender

n\_typ type: NounType idem

pron\_typ/quant\_typ feat T: PronounType/QuantifierType idem

n\_feat\_s (co,pro) feat {}: traits nominaux, NounFeature idem

?pas de structure d’arguments

**Adverbe**

adv\_typ type: AdverbType idem

adv\_val subtype: valeur de l’adverbe, Value idem

**Adjectif**

déclinaison (declension) case (strong=10, weak=11, mixed=12)

adj\_inflexion inflexion: regular,exceptional ?inflexion: AdjectiveInflexion

adj\_feat\_s feat{}: AdjectiveFeature idem

**Verbe**

v\_inflection inflexion: regular,exceptional ?inflexion: VerbInflexion

v\_typ type: VerbType idem

aux\_typ/mod\_typ feat T idem

v\_feat\_s (ordinary/causative:) feat ?feat(VerbSynFeature)fea2(VerbSemFeature)

prtl\_table feat2, feat3 subtype: particule du verbe, Particle

**arguments**

gf a1\_gf: GrammaticalFunction idem

theta a1\_theta: ThematicFunction idem

subj, dir\_obj, indir\_obj:

np\_selection a1\_sel\_feat {}: Case T idem

prep\_typ type: 99 type: -1

pp, pp2:

prep\_selection a1\_sel\_feat {}: BasicPreposition T idem

prep\_typ type: PrepositionType idem

prep\_case type + 1000  cas sélectionné (1 seul) : Case idem

s\_obj:

S\_selection a1\_sel\_feat {}: SFeature idem

pred\_object:

pred\_selection a1\_sel\_feat {}: PredFeature idem

**Allemand**

...

\begin{description}

\item[P] WordFeature uniquement :

\begin{enumerate}

feat} (type de préposition)

GBConst.time\\_p..other\\_p(9)\\

(flag particule) GBConst.particle(10)

type} (préposition de base)\\

GBConst.p\\_in..p\\_other(15)

number} (cas sélectionnés)

GBConst.nominative..distributive(5)

\end{enumerate}

\item[N] WordFeature :

\begin{enumerate}

number} (inflection nominale)

\end{enumerate}

LexemeFeature :

\begin{enumerate}

type} (type de N) GBConst.pronoun..quantifier(3)

feat} (genre) 29(masculine),30(feminine),31(neuter)\\

(type de pronom) GBConst.wh\\_pronoun..mixt(8)\\

(type de quantificateurs) GBConst.jederman..welcher(16)\\

(traits nominaux) GBConst.human..nom\\_compose(27)

inflexion} (inflection) GBConst.regular..exceptional(1)

\end{enumerate}

\item[V] WordFeature :

\begin{enumerate}

number} (inflection verbale)

\end{enumerate}

LexemeFeature :

\begin{enumerate}

type} (type de verbe) GBConst.auxiliary..causative(3)

feat} (traits verbaux) GBConst.ergative\\_v..generic\\_v(31)\\

(type de modal) GBConst.wollen..sollen(5)\\

(type d’auxiliaire) GBConst.sein..werden(2)

feat2} (particules) GBConst.be\\_prtl..zurueck\\_prtl(31)

source} (particules) GBConst.durch\\_ptrl-10..wider\\_prtl(38)-10\\

(nb de particules) 0..3+28

arg.nbArg} (nombre d’arguments 0..4

arg.sc[0..4].gf} (fonction grammaticale)

GBConst.subject..S\\_object(6)

arg.sc[0..4].theta} (rôle thématique)

GBConst.agent..undefined

arg.sc[0..4].status} (status) GBConst.unSaturated..absorbed(2)

arg.sc[0..4].selFeat} (traits sélectionnels NP-cas)

GBConst.nominative..distributive(5)

arg.sc[0..4].selFeat} (traits sélectionnels PP)

GBConst.p\\_in..p\\_other(15)

arg.sc[0..4].selFeat} (traits sélectionnels CP)

GBConst.konjugiert\\_S..konjunktiv\\_1\\_S(4)

arg.sc[0..4].selFeat} (traits sélectionnels FP)

GBConst.predicative\\_AP..predicative\\_DP(4)

\end{enumerate}

\item[D] WordFeatures :

\begin{enumerate}

number} (inflection det)

feat} (sélection) NP(0), NPmass(1), NPdenom(2), DP(3)

type} (type de déclinaison)

\end{enumerate}

LexemeFeatures :

\begin{enumerate}

type} (type de déterminant)

GBConst.definite\\_det..comparative\\_det(12)

feat2} (personne de référence-possessif) 1..6

\end{enumerate}

\item[Conj]\mbox{~~}

\begin{enumerate}

number} (sélection)

GBConst.multipleConjuncts..CPwhcompl(14)

type} (type de conjonction)

GBConst.coordination..subordination

feat} (valeur de la conjonction)

GBConst.v\\_comparison..v\\_frequency(22)

feat} (traits de conjonction)

GBConst.no\\_infinitival\\_conj..en\\_conj(12)

\end{enumerate}

\item[A] WordFeature :

\begin{enumerate}

number} (inflection adjectivale)

\end{enumerate}

LexemeFeature :

\begin{enumerate}

feat} (traits d’adjectif) GBConst.tough\\_adj..perfect\\_participle(10

inflexion} (inflection) GBConst.regular..exceptional(1)

arg.nbArg} (nombre d’arguments) 0..1

arg.sc[0..1].gf} (fonction grammaticale)

GBConst.subject..S\\_object(6)

arg.sc[0..1].theta} (rôle thématique)

GBConst.agent..undefined

arg.sc[0..1].status} (status) GBConst.unSaturated..absorbed(2)

arg.sc[0..1].selFeat} (traits sélectionnels NP-cas)

GBConst.nominative..distributive(5)

arg.sc[0..1].selFeat} (traits sélectionnels PP)

GBConst.p\\_in..p\\_other(15)

arg.sc[0..1].selFeat} (traits sélectionnels CP)

GBConst.konjugiert\\_S..konjunktiv\\_1\\_S(4)

arg.sc[0..1].selFeat} (traits sélectionnels FP)

GBConst.predicative\\_AP..predicative\\_DP(4)

\end{enumerate}

\item[Adv] WordFeature uniquement :

\begin{enumerate}

type} (type d’adverbe)

GBConst.adv\\_negation..adv\\_sentential(5)

feat} (valeur)

GBConst.v\\_comparison..v\\_frequency(22)

\end{enumerate}

\item[Interj]

\end{description}

\subsection{Anglais}

\subsection{Italien}

\section{Information syntaxique}

Les structures de données concernées Feature et les extensions spécifiques.

\begin{verbatim}

Feature = RECORD

projFeat : BITSET32;

flags : BITSET32; (\* temporary flags\*)

openChains : CARDINAL; (\* number of unfinished chains \*)

cat : CARDINAL;

infl : BITSET32;

gender : BITSET32;

number : BITSET32;

person : BITSET32;

tense : BITSET32;

type : CARDINAL;

feat : BITSET32;

feat2 : BITSET32;

selection : BITSET32;

extF : SYSTEM.ADDRESS; (\* language-specific ext. \*)

END;

\end{verbatim}

\subsection{Allemand}

Seule l’information projetée à partir des structures lexicales est discutée

ci-dessous.

\subsection{Français}

\subsection{Anglais}

\subsection{Italien}

\end{document}