EAGLES specifications for English morphosyntax
DRAFT VERSION

Simone Teufel

July 31, 1996

Contents

1 General structure of the guidelines

There is a section for each of the EAGLES main categories: *nouns, verbs, adjectives, adverbs, determiners, pronouns, articles, adpositions, conjunctions, numerals, interjections, unique-member, residuals*. Each section is structured internally according to the principles explained in the following.

1.1 Type hierarchy table

At the beginning of each section, there are tables showing the type hierarchy of the given main category. The tables are structures in two main columns, the right one for features, the left main column (which is structured) for the values applicable for the given feature. In the other dimension, the table is divided into two zones: Zone 1 displays the hierarchical relationships between the subcategory features, zone 2 the morphosyntactic features.

Zone 1 could be illustrated in a tree (as exemplarily shown in the noun section). The hierarchy in zone 1 leads to a number of terminal classes (3 in our example).

Zone 2 sums the morphosyntactic features applying to each of these classes. The value '—' indicates non-applicability of the feature to the whole terminal class, a * value indicates that there might be individual wordforms in the terminal class that the given value does not apply to. There is no hierarchical relation between features of zone 2; they *cross-classify* the feature space.

1.2 `pos`-Value section

Each main category section starts with a description of the central semantics of the class. `pos` is a special type of feature; as `pos` encodes our main categories, its values have to appear in different sections, in contrast to all other features. POS-Delimitation tables in the beginning of each main category section handle case where confusion with values of other classes/features is possible (heading *pos delimitation tables*).

1.3 Single Feature section

For each feature apart from `pos`, there is a single subsection for the description of the feature and all its values within the given main category section. This faciliates the use of the guidelines for manual taggers, as it is usually *one* value one is unsure about. A table explains the central semantics of each value by listing subclasses of the class and giving examples. General criteria for the choice of the treated value are described in a written section.

Then, there is a part that deals with delimitation *within* the values of one feature (tables for pairs of values "Delimitation value 1 vs. value 2"). These delimitations are treated pairwise for all combinations where confusion is possible. We deliberately chose examples which are very close to each other, in order to shed light on the exact borderline between two confusing contexts. Examples are always given in relevant context (sentences or parts thereof, where enough for understanding).

1.4 Example list

Relevant examples occurring in the single feature section are repeated at the end of each section, but with full analyses.[1]

1.5 Wordform index

An index of ambiguous or difficult wordforms in the appendix points to the pages where the tags for these wordforms are treated and naming the possible tags for the wordform. Wordforms occuring in the index should include

[1] In the single feature section, the wordforms were only used to illustrate differences wrt the given feature, whereas the example section is thought of as a collection of *complete* analyses.

- all word forms used as examples
- the most frequent or ambiguous members of closed classes (function words, pronouns, determiners ...)

1.6 Acknowledgement

These guidelines are a very rudimentory first definition of an English tagset, relying on the ELM suggestions. For language material and a quick survey of existing tagging practice, I relied on [Santorini 91] and [Sampson 95]. I owe much to Geoffrey Leech, who corrected and commented on the first draft of these guidelines.

2 General rules for all categories

- Abbreviations and initials are tagged as if they were spelled out: **S&L**/`[noun & common]` (savings and loan), the **U.S.**/`[noun & prop]`
- It is very advisable to have a handling of multi word items. A very good solution would be a tag to cover the whole expession or single tags where the position of the tagged item within the whole expression can be inferred[2]:

  - **due**/`adpos21` **to**/`adpos22`
  - **as**/`adpos21` **of**/`adpos22` 23th
  - **considering**/`conj21` **that**/`conj22`
  - **as**/`conj31` **long**/`conj32` **as**/`conj33`

  The above two multi word adpositions are both 2 items long (first digit after category tag), and *due* is the first item in the multi word unit *due to* (second digit, 2<u>1</u>), *to* is the second item, 2<u>2</u>.

  However, in these guidelines, no special tags are given to multi word items yet, for technical reasons (most of the existing tagsets and taggers don't handle multi word items) and theoretical reasons (the correct treatment of multi word items is a shared task between lexicon, tokenizer, maybe with some feedback from taggers or (partial) parsers). As long as these interrelationships are not fully described yet, we treat multi word items in the following simple fashion:

  - in short multi word units (exactly two word forms length), each component is annotated with the tag that should be given to the unit as a whole.
    - **due**/`adpos` **to**/`adpos`
    - **as**/`adpos` **of**/`adpos` 23th
    - **so**/`conj` **that**/`conj`
    - **as**/`conj` **if**/`conj`
    - **thanks**/`adpos` **to**/`adpos` Josephy's help
    - **considering**/`subord` **that**/`subord`
  - in longer multi word units, each component is tagged the same way as in its standard behaviour elsewhere in the text. For consistency, these decisions must be recorded in the guidelines (as they may be arbitrary).
  - they gave him, **last**/`adj` **but**/`conj` **not**/`neg` **least**/`adj` , a big meal.
  - **as**/`prep` **long**/`adj` **as**/`prep`
  - **in**/`adpos` **so**/`adv` **far**/`adj` **as**/`conj`

[2] cf. [Leech, Wilson 93]

3 N O U N S

3.1 Type hierarchy table

POS	pos	noun	
	type	com	prop
		sugar	*Billy*
MO/SY	num	sg + pl	

3.2 pos-value `noun`

3.2.1 Central semantics

- Common and proper nouns are annotated as nouns.
- Abreviations of nominal item are also annotated as nouns.
- Gerunds are in general not tagged as nouns but as *-ing* verbs. The distinction gerund-participle is in some cases not clear (*she kept on running*).
- Lexicalized gerunds like *warning* are tagged as nouns.
- Adjectival/verbal heads of NPs are not tagged as nouns, as they do not show the full morphosyntactic behaviour of nouns (e.g. Plural *-s*):
  - the **rich**/`adj` , the **handicapped**/`adj` .

3.2.2 `pos`-delimitation

Delimitation with adjectives

- difficult for noun modifiers:
  - a **fun**/`adj` party (cf. *the most fun party I ever went to*)
  - **income-tax**/`noun` return
  - **income**/`noun` **tax**/`noun` return
  - **value-added**/`adj` tax
  - **value**/`noun` **added**/`[verb & part & past]` tax
  - **terminal**/`noun` type
  - **terminal**/`adj` cancer
- he took a **sweet**/`noun`
- I find icecream too **sweet**/`adj`
- a nice **red**/`noun`
- not too many **reds**/`noun` go with that purple
- these plants are a dark **green**/`noun`
- the very **rich**/`adj`
- the **handicapped**/`adj`
- a lot of **good**/`noun`
- little **good**/`noun` will come of it

- could be adj or noun: **top**/`adj/noun` [3], **side**/`adj/noun` , **bottom**/`adj/noun` **front**/`adj/noun` , **back**/`adj/noun` .
- wordforms denoting members of nationalities which cannot occur with the indefinite Article are classified as `adj`:
  - the **Welsh**/`adj` (* a Welsh)
  - the **French**/`adj` (* a French)
  - the **Dutch**/`adj` (* a Dutch)
  - the **English**/`adj` (* an English)
  - the **Chinese**/`noun` (a **Chinese**/`noun`)
  - the **German**/`noun` (a **German**/`noun`)
- **English**/`adj` cuisine
- the **English**/`adj` are ...
- an **English**/`adj` sentence
- he's a **West**/`noun` **German**/`noun` .
- **non-NATO**/`noun & prop`
- **more**/`adj` of the same
- **most**/`adj` of the time

Delimitation with <u>pronouns</u>:
The following word forms function as heads of NPs, but are regarded as indefinite pronouns in ELM-EN:

- the only good **one**/`pron`
- the only good **ones**/`pron` of their kind
- one of them is good, but the **other**/`pron` is bad.
- **naught**/`pron` , **none**/`pron` , **anyone**/`pron` , **anything**/`pron` , **everyone**/`pron` , **everything**/`pron` , **nothing**/`pron`
- **no**/`det` **one**/`pron` [4]
- **no-one**/`[pron & indef]`

Delimitation with <u>adverbs</u> :

- nouns used adverbially: he visits us **Sundays**/`noun` and **holidays**/`noun`
- points of the compass: two miles to the **north**/`noun` of here, two miles **north**/`adv` of here.
- **yesterday**/`adv` **today**/`adv` **tomorrow**/`adv` (distributional reasons).
- call me when you get **home**/`adv` vs. when you are at **home**/`noun` .

[3] This is NOT a portmanteau tag, but only a graphical device to show that exactly one of the two tags `adj` or `noun` is given, depending on the context.
[4] As Geoffrey Leech pointed out, a multiword tag [pron & indf] would be more satisfactory here.

Delimitation with gerunds/present participles:

- good **cooking**/`noun` is something to enjoy
- **cooking**/`verb` well is difficult
- GM's (repeated) **closing**/`verb` of the plant
- GM's (repeatedly) **closing**/`verb` the plant
- unsavory **plant**/`noun` **closing**/`noun` **tactics**/`noun & pl`
- **mating**/`noun` season
- **spending**/`noun` reductions
- **anchoring**/`noun` devices

Delimitation with past participles:

- her **firstborn**/`noun` was feeble

3.3 Feature `ntype`

criteria:

- graphical criterion: upper case /lower case
- semantics: reference to a single entity
- syntax: lack of article $->$ prop

3.3.1 Value `prop`

- surnames
- forenames
- initials of person name
- country or province name
- geographical proper name
- product names (a **Mars**/`prop` bar)
- names of fictional/mythological figures, planets
- names of organisations, parties
- titles, forms of address[5] **Sir**/`prop` **Miss**/`prop` **Lady**/`prop`

[5] This is an exception, which contradicts the semantics and syntax criterion. It was decided, however, purely for distributional reasons.

3.3.2 Value `com`

- months, days of the week: **Sunday**/`com` , **March**/`com` , **mid-March**/`com` .
- inhabitants of places: a **Londoner**/`com` , a **Swiss**/`com` , the **English**/`com`
- concrete nouns
- abstract nouns
- (unit) measurements (**mile**/`com`)
- (non-exact) measurements (a **bag**/`com` of clothes)
- currencies (**Dollar**/`com`)

List of examples[6]

- a **Swiss**/`com`
- a **Chinese**/`com` , **Chinese**/`com` is tonal
- a **fish**/`com` , three **fish**/`com` , a piece of **fish**/`com`
- **midstream**/`com` , **midfield**/`com`
- **Corp**/`com` **Bros**/`com`
- **Hallowe'en**/`com`
- **noon**/`com` **midnight**/`com` **midyear**/`com`
- I fed Aunt Louise's **Airedale**/`com`

3.4 Feature `num`

criteria:

- First criterion[7]: agreement within the NP (determiners, numerals)
- Minor criterion: verb-subject agreement
- Semantics is not a criterion

3.4.1 Value `sg`

- the **crowd**/`sg` were arguing.
- **linguistics**/`sg` is a difficult field
- the **group**/`sg`
- the **jury**/`sg`
- the **faculty**/`sg` are on strike
- **people**/`pl` say (*people* is regarded as an irregular plural form of *person*).
- a **person**/`sg`
- a Central African **people**/`sg`
- Central African **peoples**/`pl` and **tribes**/`pl`

[6] Most of the examples were taken from [Sampson 95], and converted into our tagset. ELM-EN follows Geoffrey Sampson's suggestions in these examples.

[7] Personal communication with Geoffrey Leech.

3.4.2 Value `pl`

- **scissors**/`pl`
- but: three **years**/`pl` is a long time
- 12 **inches**/`pl` is a foot
- the **Messrs.**/`pl` are coming

4 V E R B S

4.1 Type hierarchy table

POS	pos	verb			
	vform	fin			
	vtype	aux	modal	mai	s_aux
		are	*would*	*cooks*	*ought*
MO/ SY	tense	pres + past			—
	num	sg + pl			—
	pers	1+2+3			—
	vf_m	ind + impr		ind	—

POS	pos	verb					
	vform	no_fin					
	vf_m	inf		papa		ing	
	vtype	aux	mai	aux	mai	aux	mai
		have	*boil*	*done*	*beaten*	*being*	*going*
MO/SY	—						

4.2 pos-value `verb`

4.2.1 Central semantics

questions:

- No distinction between gerunds, present participles is drawn. Present participles either
  - used with to be (present continous, belonging to verbal complex)
  - used as a head of a subclause

  are treated alike with gerunds.
- We distinguish finite non-marked forms from infinitives.
  - How **do**/`no_fin` they **do**/`finite` ?

4.2.2 `pos`-delimitation

Delimitation gerund/present participle with nouns:
cf. section 3.2.2

Delimitation past participles with adjectives:

- a **mustachioed**/`adj` policeman
- he was more **surprised**/`adj` than she was (gradability)
- an **unhurried**/`adj` meeting (un-prefixability)
- Your shoelace has been **untied**/`verb` ever since we started
- it got **untied**/`verb` by accident
- **armed**/`adj` guard

- **armed**/`verb` with only a knife...
- he became **interested**/`adj`
- he remains **guided**/`verb` by these principles (because of by-phrase)
- should be kept well **watered**/`verb`
- at the time, I was **married**/`adj` (state)
- I was **mistaken**/`verb` the other day
- a **mistaken**/`adj` decision
- a case of **mistaken**/`adj` identity
- **decided**/`adj` advantage (can't be passivized)
- **grown**/`adj` woman
- **married**/`adj` life
- **worried**/`adj` faces
- I was **invited**/`verb` by friends (by-phrase)
- **surprised**/`verb` by her remarks
- I was **married**/`verb` on a Sunday (replacable by *get*, not by *become*)
- **surprised**/`adj` at the news
- **surprised**/`verb` by sudden rain falls
- **pleased**/`adj` with the day's start
- **pleased**/`verb` by too many admirers
- **worried**/`adj` about his progress
- **worried**/`verb` by her remarks
- **involved**/`adj` in discussions
- **involved**/`adj` with these problems
- got **involved**/`verb` by mysterious circumstances

Delimitation present participles with adjectives:

- gradability: her talk was more **interesting**/`adj` than his.
- an **uninteresting**/`adj` conversation
- that place feels **depressing**/`adj`
- cannot be turned into a relative clause: an **appealing**/`adj` face
- an **appetizing**/`adj` dish
- a **revolving**/`adj` fund

- a **winning**/`adj` smile
- can be turned into relative: the **existing**/`verb` safeguards
- a **managing**/`verb` director
- the **ruling**/`verb` class
- a **striking**/`adj` hat
- the **striking**/`adj` teachers (can be turned into relative clause)[8]
- a **thoroughgoing**/`adj` examination (* *to thoroughgo*)
- **outgoing**/`adj` type of guy
- **outstanding**/`adj` quality
- equivalence to a relative clause *which is X.ing*:
  - **declining**/`adj` productivity of U.S. industry
  - the **acting**/`adj` vice president

Delimitation past participles with prepositions:

- **provided**/`verb` that[9]
- **granted**/`verb` that
- **given**/`verb` fine weather

4.3 Feature `vform`

The distinction between finite and non-finite is drawn along the following lines:

- Distinction between Present non-3rdSg forms and infinitive forms;
- distinction between past participle and finite past tense verbs forms;
- subjunctive forms are treated as finite verbs;
- imperatives are treated as infinite verbs.

[8] Geoffrey Leech drew my attention to the fact that there would be conflicting criteria for the example *an* **employed**/`adj??verb` *teacher* if we decided to annotate *striking* in *striking teachers* as `verb`:

- *an unemployed teacher* $->$ `adj`
- *an employed teacher* = a teacher that is employed by ... $->$ `verb`

[9] A multiword tag would be the method of choice here and in the next example.

4.3.1 Value `fin`

- non-inflected verb forms, acting as finite verb in sentence:
  - they **do**/`fin` not **understand**/`no_fin` .
  - they **understand**/`no_fin` everything.
- inflected finite forms
  - he **speaks**/`fin` Greek well.
  - I **am**/`fin` Greek myself.
  - finite past tense forms
    * They **finished**/`fin` the reading late in the afternoon.
    * We **had**/`fin` **finished**/`no_fin` the drawing by then.

4.3.2 Value `no_fin`

- ing-forms;
- infinitives;
- participles perfect;
- (non-inflected) subjunctive forms (where distinguishable, that is in 3rd Person Sg and with some special verbs).
  - The rule demands that he **be**/`no_fin` on time.
  - He demands that he **shout**/`no_fin` no more.
  - I recommended that we **be**/`no_fin` on time.

4.4 Feature `type`

No distinction between auxiliary and main verb uses of *be, have, do* is introduced yet (technically difficult). This distinction is desirable for the future, however.

4.4.1 Value `aux`

This value comprises primary and modal auxilaries.

4.4.2 Value `s_aux`

- **able**/`s_aux` [to]
- **ought**/`s_aux` [to]
- **bound**/`s_aux` [to]
- **going**/`s_aux` [to]

4.4.3 Value `mai`

Main verbs (or content verbs), other than the ones in `aux`. Classification is based on the lemma, not on the function.

4.5 Feature `aux-type`

4.5.1 Value `prim`

Primary auxiliaries are only *be*, *have* and all forms thereof.

4.5.2 Value `modal`

Modal auxiliaries are *can, will (would), may (might), shall (should)*

4.6 Feature `tense`

This feature applies only to finite verbs. There should not be any problems. There are ambiguous forms like *cut* that have the same appearance in Present Tense and Past Tense. They must be disambiguated by context.

4.6.1 Value `pres`

- I never **cut**/`pres` my own hair

4.6.2 Value `past`

- ... but yesterday, I just had to: I **cut**/`past` my hair as my hairdresser was on holiday.

4.7 Feature `num`

Most important criterion: subject-verb agreement.

4.7.1 Value `sg`

- this sheep **runs**/`sg`
- the crowd **is**/`sg` cheering

4.7.2 Value `pl`

- the police **are**/`pl`
- the sheep **run**/`pl`

4.8 Feature `pers`

4.8.1 Value `1`

This feature is not handled in the tagging task documented in this report.

4.8.2 Value `2`

4.8.3 Value `3`

4.9 Feature `vf_m`

4.9.1 Value `ind`

For finite forms

4.9.2 Value `impr`

For finite forms: imperatives

4.9.3 Value `inf`

Under this value, we subsume real infinitives and subjunctive forms.

4.9.4 Value `papa`

past participles

4.9.5 Value `ing`

Present participles or gerunds. The distinction between the two is not drawn: it is too difficult to draw (and as well, it is controversial if the distinction is useful at all).

Example list[10]

- **clad**/`[verb & papa]`
- the **observed**/`[verb & papa]` data
- many **sitting**/`adj` targets
- the headfirst **observed**/`papa` in the field
- the above **mentioned**/`adj` bodies
- the **aged**/`adj` and the **maimed**/`adj`
- or a **charged**/`adj` particle will stick to an **uncharged**/`adj` surface
- **exposed**/`adj`
- **unprotected**/`adj`
- **unexpected**/`adj`
- **downtrodden**/`adj`
- her **firstborn**/`noun` was feeble
- the **aged**/`adj`

[10] Material adapted from [Sampson 95].

5 A D J E C T I V E S

5.1 Type hierarchy table

POS	pos	adj
		joyful
MO/SY	degr	po + comp + sup

5.2 pos-value `adj`

5.2.1 Central semantics

- The stereotypical distribution of adjectives is between the article and head noun (if there are nominal modifiers: before those).
- Adjectives may be gradable
- Ordinal numbers are not tagged as adjectives but as a special class `[number & ordinal]`, as they have no morphology and can occur in adjectival, nominal and adverbial function.
- adjectival wordforms used as heads of NPs, like the *rich*, the *poor* are tagged as adjectives (morphological criterion).
- wordforms denoting members of nationalities which cannot occur with the definite Article are classified as `adj` (cf. section 3.2.2 above: the **Welsh**/`adj`
- adjectives that can be used only attributively[11] are:
  - **chief**/`adj`
  - **entire**/`adj`
  - **future** /`adj`
  - **inverse**/`adj`
  - **major**/`adj`
  - **planetary**/`adj`
  - **so-called**/`adj`
  - **undersea**/`adj`
  - the **very**/`adj` thing
- adjectives that can be used only predicatively are:
  - **ablaze**/`adj`
  - **alone**/`adj`
  - **unable**/`adj`
- complex adjectives: **one-of-a-kind**/`adj` , **fourth-largest**/`adj`

5.2.2 `pos`-delimitation

Delimitation with <u>verbs</u> (past and present participles)
cf. section 4.2.2

[11] This list is taken from [Sampson 95]

Delimitation with <u>nouns</u>:
cf. section 3.2.2

Delimitation with <u>cardinal numbers</u>:

- a **50-3**/`adj` victory
- **one**/`card` **half**/`adj` cup
- nearly a **quarter**/`noun` the amount
- **one**/`card` **half**/`noun` the amount (measure phrase). The noun annotation in this and the previous example is justified by very similar nominal constructions: cf.
  - 3 **times**/`noun` the amount
  - several **blocks**/`noun` north of here

Delimitation with <u>adverbs</u>:

- **rapid**/`adj` growth
- **rapid**/`adj` **growing**/`[verb & ing]` plants
- **quick-growing**/`adj` city
- **slow-spreading**/`adj` disease
- **widespread**/`adj` believes
- **wide-ranging**/`adj` thoughts
- **mild**/`adj` **flavored**/`adj` chips
- make life **simple**/`adj`
- **utmost**/`adj` danger
- **northernmost**/`adj` town
- **outermost**/`adj` planet
- shares closed **higher**/`adj` (shares were higher when they closed, not: shares closed more highly).
- shares closed more **reasonably**/`adv` today
- shares closed **up**/`[adv & particle]` by two points

Delimitation with <u>pronouns/determiners</u>:

- **such**/`det` pleasure!
- **such**/`det` a sight
- There are many **such**/`pron`

5.3 Feature `degr`

criterion: morphology, not semantics

5.3.1 Value `po`

- **top**/`po`

5.3.2 Value `comp`

- **elder**/`comp`
- **further**/`comp`

5.3.3 Value `sup`

6 P R O N O U N S

6.1 Type hierarchy table

POS	pos	pron						
	type	personal	refl	poss	dem	indf	wh	
	wh_type	—	—	—	—	—	rel	interr
		she	*themselves*	*mine*	*this*	*some*	*whom*	*which*
MO/ SY	pers	1 + 2 + 3	1 + 2 + 3	—	—	—	—	—
	case	obl + nom	—	—	—	—	nom + gen + obl	
	num	sg + pl	sg + pl	—	—	—	—	—
	gend	m + f + n	m + f + n	—	—	—	—	—

6.2 pos-value `pron`

6.2.1 Central semantics

- Criteria for pronouns:
  - stand-alone as NPs:
    - **many**/`pron` came, **we**/`pron` came
  - heads of partitive constructions:
    - **much**/`[pron & indef]` of his talk
    - **many**/`[pron & indef]` of these
- wh-pronouns can be subcategorised into relative, interrogative and exclamatory wh-pronouns.
- Personal, reflexive and possessive pronouns are distinguished.
- "postdeterminers" are annotated as pronouns:
  - they **both**/`[pron & indef]` came to the party
  - the girls **each**/`[pron & indef]` wanted a cake.
- **Each**/`[pron & indef]` /**Both**/`[pron & indef]` of them came
- "Antecedentless relatives" (*I eat* **what**/`[pron & wh]` *you eat*) are annotated as relatives. The distinction in these cases between relative and interrogative pronoun is nontrivial.
- *where, when, how, why...* are regarded as wh-adverbs, not as interrogative pronouns.

6.2.2 `pos`-delimitation

Delimitation with <u>determiners</u>:

- syntactic properties: pronouns are heads of NPs
  - **Many**/`det` party-goers prefer wine to beer.
  - **Many**/`pron` disagreed to the leader's speech.
- "postdeterminer" position $->$ pronuns:
  - the girls **both**/`pron` liked cake.
  - **both**/`det` the girls went swimming.

- they **both**/pron gave good answers
- I'll take **both**/pron .
- the employees have **all**/pron left for a holiday.
- **all**/det the food is gone.
- The girls **each**/pron took a sweet.

Delimitation with adverbs:

- Her hair was **all**/adv (= completely) in a mess.
- I am **all**/adv (= totally) upset.
- The cakes have **all**/pron (= each of them) been eaten.

6.3 Feature type

6.3.1 Value personal

I/personal , **me**/personal , **you**/personal , **he**/personal , **him**/personal ...

6.3.2 Value refl

*-self, *-selves: **myself**/refl , **ourselves**/refl ...

6.3.3 Value poss

Only those that act as the head of an NP: This is **mine**/poss , The hat was **hers**/poss , **yours**/poss , **ours**/poss , **theirs**/poss ...

6.3.4 Value dem

Only where they act as heads of nps: *So what about* **these**/dem *?*
that/dem , **these**/dem , **those**/dem , **this**/dem

6.3.5 Value indf

- one as in *the green* **one**/indf
- **some**/indf , **many**/indf , **little**/indf , **more**/indf , **most**/indf , **several**/indf **all**/indf , **each**/indf ...as head of NP.
- the good **ones**/[indf & pron]
- { **some, any, no**}{**one, body**}/indf
- **less**/indf , **least**/indf

6.3.6 Value wh

- I never understood **what**/wh he said.
- **What**/wh did she say again?

6.4 Feature `wh_type`

6.4.1 Value `rel`

- Relative pronouns that act as a NP of their own:
  - the man **whom**/`[rel & pron]` I met
  - the guy **who**/`[rel & pron]` I never recognized
  - the child **who**/`[rel & pron]` got kidnapped
  - the guy **that**/`[rel & pron]` I never recognized
- antecedentless relatives (Sampson):
  - I cannot comment on **what**/`rel` I have never seen.
  - I vote for **whoever**/`rel` promises the most.
  - You can take **whichever**/`rel` you like best.

6.4.2 Value `interrog`

- In direct questions:
  - **Who**/`[interrog & pron]` arrived yesterday?
  - **What**/`[interrog & pron]` was the hassle all about?
  - **Whom**/`[interrog & pron]` did you meet?
- in indirect questions:
  - He wondered **what**/`[interr & pron]` people worried about
  - **Whom**/`[interr & pron]` I gave the book to, everybody asked me

6.5 Feature `pers`

6.5.1 Value `1`

6.5.2 Value `2`

6.5.3 Value `3`

6.6 Feature `case`

6.6.1 Value `nom`

6.6.2 Value `gen`

6.6.3 Value `obl`

6.7 Feature `num`

6.7.1 Value `sg`

6.7.2 Value `pl`

6.8 Feature `gend`

6.8.1 Value `masc`

6.8.2 Value `fem`

6.8.3 Value `neutr`

7 D E T E R M I N E R S

7.1 Type hierarchy table

POS	pos	det					
	type	poss	dem	indf	wh		
	wh_type	—	—	—	rel	excl	interrog
		my	*this*	*some*	*whose*	*which*	*what*
MO/	pers	1 + 2 + 3	—	—	—	—	—
	num	—	sg + pl	sg + pl	—	—	—
	n_pos	sg_p + pl_p	—	—	—	—	—
	gend	m + f + n	—	—	—	—	—

7.2 pos-value `det`

7.2.1 Central semantics

- Criteria for determiner position: Noun following
  - **this**/`det` problem
  - **their**/`det` suggestion
  - **much**/`det` ado
- Predeterminer position: before (definite or indefinite) article and noun:
  - adverbial:
    - **rather**/`adv` a bad day
    - **quite**/`adv` a mess
  - indefinite determiners
    - **such**/`[det & indf]` a good time
    - **both**/`[det & indef]` the children
    - **all**/`[det & indef]` those years
  - wh determiners
    - **what**/`[det & wh]` a fool!
- archaic/dialected demonstrative determiners:
  - *yonder*:
    - in **yonder**/`[det & demo ]` field
    - BUT: they came galloping over that hill **yonder**/`adv` .
  - *yon*: What's **yon**/`[det & demo ]` place in America, Coney Island?

7.2.2 `pos`-delimitation

Delimitation with <u>determiners</u>:
The criterion is syntax (distribution). Pronouns act as heads of NPs or are in "postdeterminer" position. Cf. section 6.2.2.

7.3 Feature `type`

7.3.1 Value `poss`

- **my**/`poss` , **your**/`poss` , **his**/`poss` , **her**/`poss` , **its**/`poss` , **our**/`poss` , **their**/`poss` , **one's**/`poss`

7.3.2 Value `dem`

- **this**/`dem` car, **that**/`dem` question
- **those**/`dem` problems, **these**/`dem` cars
- **yon**/`dem` place, **yonder**/`dem` forest

7.3.3 Value `indf`

- determiner position
  - **both**/`[det & indf]` daughters
  - **all**/`[det & indf]` roads
  - **another**/`[det & indf]` person
  - **any**/`[det & indf]` problems
  - **some**/`[det & indf]` sugar
  - **each**/`[det & indf]` parent
  - **either**/`[det & indf]` way
  - **no**/`[det & indf]` chance
  - **neither**/`[det & indf]` decision
  - **every**/`[det & indf]` day
  - **several**/`[det & indf]` months
  - **few**/`[det & indf]` children
  - **little**/`[det & indf]` confusion
  - **more**/`[det & indf]` attention
  - **most**/`[det & indf]` children
  - ...
- predeterminer position
  - **such**/`[det & indf]` a good time
  - **both**/`[det & indef]` the children
  - **all**/`[det & indef]` those years
  - **many**/`[det & indef]` a moon
  - **nary**/`[det & indef]` a soul, with **nary**/`[det & indef]` a word

7.3.4 Value `wh`

Exclamatory, interrogative and relative determiners (i.e. the head of the NP is following!), as in:

- **What**/`wh` a lovely colour!
- **Whose**/`wh` car did you borrow?
- I wondered **which**/`wh` library was your favourite.
- Carl, **whose**/`wh` car I borrowed that day, took the motorbike.

7.4 Feature `wh_type`

7.4.1 Value `interrog`

Interrogative determiners that need a noun to complete an NP:

- **Whose**/`wh` car did you borrow?
- **Which**/`wh` problems did you discuss?

7.4.2 Value `rel`

Similarly to interrogative determiners, the relative determiner *whose* needs a noun to complete an NP.

- the man **whose**/`wh` daughter became ill

7.4.3 Value `excl`

- What beautiful eyes!

7.5 Feature `pers`

7.5.1 Value `1`

7.5.2 Value `2`

7.5.3 Value `3`

7.6 Feature `num`

Number applies to demonstratives and some indefinite determiners:

- **these**/`pl` **geese**/`pl`
- **this**/`sg` **sheep**/`sg`
- **many**/`pl` **children**/`pl`

The situation is a bit different for possessives:

- Number of possessors `sg`:
  - Number of possessions `sg`: **my**/`sg-sg` **book**/`sg`
  - Number of possessions `pl`: **my**/`pl-sg` **books**/`pl`
- Number of possessors `pl`:
  - Number of possessions `sg`: **their**/`pl-sg` **book**/`sg`
  - Number of possessions `pl`: **their**/`pl-pl` **books**/`pl`

Number of possessors is encoded as an extra feature for possessives (feature `n_pos`, lemma based). Number of possessions (feature `num`) is interesting, as it is the same noun semantics as with indefinite determiners etc., but not easy to annotate in an automatic procedure and *not related to the lexical lemma of possessives themselves in English*, as the above examples show. We therefore decided <u>not</u> to annotate number of possessions (`num`) for possessive determiners.

7.6.1 Value sg

7.6.2 Value pl

7.7 Feature n_pos

As explained above, this feature only applies to possessive determiners. It is a lexical feature, i.e. it does not depend on the context.

- **my**/n_pos = sg_p **book**/num = sg
- **my**/n_pos = sg_p **books**/num = pl
- **their**/n_pos = pl_p **book**/num = sg
- **their**/n_pos = pl_p **books**/num = pl

7.7.1 Value sg_p

This value means singular possessor (cf. above).

7.7.2 Value pl_p

This value means plural possessor (cf. above).

7.8 Feature gend

7.8.1 Value m

7.8.2 Value f

7.8.3 Value n

8 A R T I C L E S

8.1 Type hierarchy table

POS	pos	art	
	defin	idef	def
		an	*the*
MO/SY	num	sg	sg + pl

8.2 pos-value `art`

8.2.1 Central semantics

Simple: just the forms *the, a, an*.

8.2.2 `pos`-delimitation

8.3 Feature `defin`

8.3.1 Value `idef`

8.3.2 Value `def`

8.4 Feature `num`

8.4.1 Value `sg`

8.4.2 Value `pl`

8.5 Example list

9 A D V E R B S

9.1 Type hierarchy table

POS	pos	adv				
	polarity	wh_adv		non_wh_adv		
	type	—		general	degree	particle
	wh-type	rel	interrog	—	—	—
		where	*how*	*awfully*	*very*	*up*
MO/SY	degr	—		po + comp + sup	—	—

9.2 pos-value `adv`

9.2.1 Central semantics

- modification of a verb:
  - the car moved **slowly**/`adv` .
- modification of an adjective phrase:
  - He was **fully**/`adv` aware of this fact
- modification of another adverb:
  - She walked **very**/`adv` slowly into the parlour.
- wh-adverbs in clause/subclause initial position:
  - **When**/`adv` are you coming?
  - I remember **when**/`adv` she was a young girl
  - **How**/`n` ice of you to come!
- Adverbs as posthead modifiers of adjectival or adverbial phrases:
  - good **enough**/`adv`
  - very well **indeed**/`adv`

9.2.2 `pos`-delimitation

Delimitation with conjunctions:

- *When* as a temporal or conditional conjunction is tagged as conjunction. Adverbial contexts are tagged as interrogative adverb, however.
  - **When**/`[conj]` he finally arrived, I was on my way out. (temporal)
  - I like it **when**/`[conj]` you make dinner for me. (conditional)
  - BUT: I remember **when**/`adv` she arrived (adverbial)

Delimitation with pronouns:

- “personal” relative/interrogative pronouns as *who, what, whom* are tagged as pronouns.

Ambiguity with adjectives:

- general adjectives[12]: *forward, nationwide, offside, overland, right, underground, up_till, ...*
- purely predicative adjectives: *away, back, backstage, downriver, eastward, inland, near, westward, overhead, offstage, sidelong...*

Ambiguity with prepositions:

- aboard[12], between, beyond, around, astride, below, by, within, inside, near, opposite, outside, past, round, throughout, under.

9.3 Feature `polarity`

9.3.1 Value `wh_adv`

Wh-adverbs are: *when, why, how...*, in all positions (in direct and indirect questions).

- I wonder when/[adv & wh_adv] we should leave.
- How/[adv & wh_adv] did you do this?

9.3.2 Value `non_wh_adv`

9.4 Feature `advtype`

9.4.1 Value `general`

- all adjectival word forms ending in *-ly*
- local adverbs
  - **here**/`general`
  - **over**/`adv & general` **there**/`adv & general`
- temporal adverbs
  - **soon**/`adv & general` , **late**/`adv & general` , **early**/`adv & general` , **daily**/`adv & general` , **always**/`adv & general` , **again**/`adv & general` , **then**/`adv & general` , **now**/`adv & general` ,...
- **far**/`adv & general` , **long**/`adv & general` , **otherwise**/`adv & general` , **only**/`adv & general` , **fast**/`adv & general`
- **elsewhere**/`adv & general` , **somewhere**/`adv & general` , **as**/`adv & general` , **so**/`adv & general` , **rather**/`adv & general` , **abroad**/`adv & general` , **afar**/`adv & general` , **ahead**/`adv & general` , **away**/`adv & general` , **back**/`adv & general` , **behind**/`adv & general` , **below**/`adv & general` , **beneath**/`adv & general` , **beside**/`adv & general` , **to**/`adv & general` , **together**/`adv & general` , **nationwide**/`adv & general` , **inland**/`adv & general` , **inshore**/`adv & general` , **forward**/`adv & general` , **forwards**/`adv & general` , **sideways**/`adv & general` , **round**/`adv & general` ...

[12] List taken from [Sampson 95].

9.5 Feature `wh_type`

9.5.1 Value `other`

This value includes interrogative and exclamatory readings, as well as “other” readings (those that are difficult to decide):

- **How**/`other` old you are! (exclamatory)
- **How**/`other` old is she? (interrogative)
- I was surprised **how**/`other` old she was (difficult to decide)

9.5.2 Value `rel`

This value is for clear relative cases as in:

- The place **where**/`rel` I met you.
- The reason **why**/`rel` I did it.

9.5.3 Value `degree`

- very, hardly,

9.5.4 Value `particle`

Verbal particles (occurring in the same sentence as the respective phrasal verb).

- **up**/`particle` , **off**/`particle` , **on**/`particle` , **to**/`particle` , **down**/`particle` , **in**/`particle` , **out**/`particle` , **off**/`particle` , **on**/`particle` ...

The distinction with prepositions is difficult for automatic tagging.

9.6 Feature `degr`

Morphology decides, not semantics!

9.6.1 Value `po`

- we can always come by **later**/`comp` (comparative meaning missing; however, morphosyntactic criterion).
- **sooner**/`comp` or **later**/`comp`

9.6.2 Value `comp`

9.6.3 Value `sup`

10 A D P O S I T I O N S

10.1 Type hierarchy table

POS	pos	adpos	
	adpos-type	preposition	postposition
		off	*'s*
MO/SY	—		

10.2 pos-value `adpos`

10.2.1 Central semantics

- Complex adpositions (multi word units):[13]
  - **due**/`adpos` **to**/`adpos`
  - **as**/`adpos` **of**/`adpos` 23th of June
  - **thanks**/`adpos` **to**/`adpos` Josephy's help

10.2.2 pos-delimitation

- difficult to distinguish from verbal particles
  - turn **off**/`[particle]` the radio
  - swim **off**/`[adpos]` the shore

  As the examples show, this distinction can be purely semantic: In the first sentence, the direct object is a machine, which makes *turn off* a phrasal verb. In the second sentence, syntactically identical, the direct object is a geographic place, which makes *off the shore* a local prepositional phrase. To do this distincition automatically is nearly impossible: Apart from a comprehensive lexicon of phrasal verbs, semantic information is needed.

- difficult to distinguish from subordinating conjunctions:
  - I went there **to**/`[conj & subord]` see if he had good news.
  - I went **to**/`[adpos]` the doctor
  - I dropped by **before**/`adpos` dinner
  - I got out **before**/`conj` they could see me
  - They met me **after**/`adpos` lunch
  - He was arrested **after**/`conj` he had injured 2 persons
  - Don't come here **until**/`adpos` 2!
  - Don't come here **until**/`conj` you have the money!
  - He was suspected **for**/`adpos` obvious reasons
  - He was suspected **for**/`conj` he had behaved stangely all
  - She was driving **like**/`s` he had not a minute to lose
  - She was driving **like**/`a` madwoman. day.

[13] As described in section 2, we are aware of the shortcomings of this simple way of tagging, but we have our reasons not to recommend special treatment of multi word items at this stage.

10.3 Feature `adpos-type`

10.3.1 Value `preposition`

- **to**/`preposition` **from**/`preposition` **of**/`preposition` **until**/`preposition` **after**/`preposition` **before**/`preposition` **like**/`preposition` **aboard**/`preposition` **about**/`preposition` **against**/`preposition` **during**/`preposition` **due**/`preposition` **down**/`preposition` **despite**/`preposition` **nearest**/`preposition` **off**/`preposition` **nearer**/`preposition` **near**/`preposition` **inside**/`preposition` **into**/`preposition` **inter**/`preposition` **less**/`preposition` **on**/`preposition` **past**/`preposition` **plus**/`preposition` **round**/`preposition` **through**/`preposition` **under**/`preposition` **underneath**/`preposition` **unlike**/`preposition` **toward**/`preposition` **pro**/`preposition` **regarding**/`preposition` **respecting**/`preposition` **wanting**/`preposition` **within**/`preposition` **worth**/`preposition` ...

10.3.2 Value `postposition`

Possessive endings **'**/`postposition` **'s**/`postposition` **s'**/`postposition` are regarded as enclitic postpositions, following Geoffrey Leech's analysis in [Monachini, Calzolari 94]:

- the **teachers**/`noun` **'**/`postposition` room
- **Carl**/`noun & prop` **'s**/`v` an
- in a month or two**'s**/`postposition` time

11 C O N J U N C T I O N S

11.1 Type hierarchy table

POS	pos	conj					
	conjtype	coord			subord		
	coordtype	simple	init	no_init	—		
	subordtype	—			s_infve	compara	s_fin
		and	*neither*	*nor*	*in order to*	*than*	*whereas*
MO/SY	—						

11.2 pos-value `conj`

11.2.1 Central semantics

Criteria:

- introductory element of a subclause
  - It will become cold **if**/`conj` you leave the door open
  - Even **though**/`conj` she is never on time, she hasn't missed a single train yet.
- introductory element of a structure capable of being expanded into a subclause:
  - They are dangerous **if**/`conj` hit
  - **Though**/`conj` a coward, she fought hard.

11.2.2 `pos`-delimitation

Delimitation with particles:
cf. adverb section.

Delimitation with prepositions:

- comparative conjunctions vs. prepositions:
  - I am larger **than**/`adpos` you. (NP following *than*)
  - I am larger **than**/`conj` you will ever be (clause following)

  There are contexts where the expansion into a clause is not possible:

  - children younger **than**/`adpos` 18.

  This fact justifies the classification of *than* as a preposition.

- prepositions *of/without* + ing-forms are not regared as conjunctions:
  - gerund subcategorized:
    without/`adpos` ever mentioning it.
  - subject + gerund subcategorized:
    - **without**/`adpos` anybody ever mentioning it again.
    - the possibility **of**/`adpos` them being killed.
  - problematic case:
    she was regarded **as**/`conj?adpos` having too much power

- *for* as a preposition vs. *for* in the sense of *because* (conjunction):
  - I asked **for**/`preposition` more money.
  - He was looking for a flat **for**/`conj` he moved out of his old one.

Delimitation with determiners/pronouns:

- **both**/`conj & coord` the girl **and**/`conj & coord` the boy
- **both**/`det` girls went
- the girls **both**/`pron` went
- **either**/`conj & coord` Suzie **or**/`conj & coord` Tom
- **either**/`det` question
- **neither**/`conj & coord` Suzie **nor**/`conj & coord` Tom
- **neither**/`det` question

11.3 Feature `conjtype`

There are two criteria to distinguish subordinating conjunctions from coodinating conjunctions:

- Subclauses introduced by a subordinating conjunction can be put into sentence initial position:
  - **As**/`subord` he was unhappy, he asked to be transferred
  - **Because**/`subord` he was unhappy, he asked to be transferred
  - *For he was unhappy, he asked to be transferred
- Subclauses introduced by a subordinating conjunction can be preceded by a coordinator:
  - He asked to be transferred **because**/`subord` he was unhappy and **because**/`subord` he had to commute.
  - * He asked to be transferred for he was unhappy and for he had to commute.

  Therefore, *for* is a coordinating conjunction.

11.3.1 Value `coord`

- **and**/`coord` , **but**/`coord` , **nor**/`coord` , **or**/`coord`
- this book is [cheap] **yet**/`coord` [good]
- mathematical operators (spelled out): **plus**/`coord` , **minus**/`coord` , **less**/`coord` , **times**/`coord` , **over**/`coord`
- *for* in the sense of *because*:

  He asked to be transferred, **for**/`[conj & coord]` he wasn't happy.

11.3.2 Value subord

- **so**/subord in the sense of *so that.*
- **albeit**/subord [14], **although**/subord , **because**/subord , **even**/subord **if**/subord , **even**/subord **though**/subord , **except**/subord **that**/subord , **as**/prep **long**/adj **as**/prep , **considering**/subord **that**/subord , **in**/subord **case**/subord , **in**/adpos **order**/noun **that**/conj , **in**/adpos **so**/adv **far**/adj **as**/conj , **in**/subord **that**/subord , **lest**/subord , **now**/subord , **once**/subord , **other**/subord **than**/subord , **provided**/subord , **provided**/subord **that**/subord , **providing**/subord , **providing**/subord **that**/subord , **seeing**/subord , **seeing**/subord **that**/subord , **unless**/subord , **whereas**/subord , **whereat**/subord , **whereupon**/subord , **while**/subord , **whilst**/subord ...

11.4 Feature coordtype

11.4.1 Value simple

Coordinating conjunctions consisting only of one lexical item:

- **and**/simple
- **or**/simple

11.4.2 Value init

Initial part of complex coordinating conjunction.

- **both**/[conj & coord & initial] the girl and the boy
- **either**/[conj & coord & initial] Suzie or Tom
- **neither**/[conj & coord & initial] Suzie nor Tom

11.4.3 Value no_init

Second, third or any other repeated part of complex coordinating conjunction. Of course, for items that are ambiguous to simple coordinating conjunctions (*and, or*), automatic disambiguation is very difficult.

- both the girl **and**/[conj & coord & noninitial] the boy
- either Suzie **or**/[conj & coord & noninitial] Tom
- neither Suzie **nor**/[conj & coord & noninitial] Tom

11.5 Feature subordtype

11.5.1 Value s_infve

- They meet at seven in order **to**/s_infve have dinner before the meeting.

11.5.2 Value compara

- I am faster **than**/compara you will ever be.

11.5.3 Value s_fin

- I hope **that**/s_fin you will recover soon.

[14]This list is again taken from [Sampson 95]

12 N U M E R A L S

12.1 Type hierarchy table

POS	pos	numeral	
	type	ord	card
		fifth	*seven*
MO/SY	number	sg + pl	sg + pl

12.2 pos-value `number`

12.2.1 Central semantics

- Numbers spelled out and written in digits are treated alike.
- Ordinal and cardinal numbers are treated under the heading "numeral", and the distinction between them is drawn.

12.2.2 `pos`-delimitation

Delimitation with <u>nouns</u>:

- *The big* **8**/`numeral` *is painted on the wall.*
- **Eight**/`numeral` horses are for sale.

Delimitation with <u>indefinite pronouns</u>:

- *one* is an indefinite pronoun when it is personal (one, oneself, one's) or when it could be replaced by ones.
  - The best **ones**/`pron` die young.
  - **One**/`pron` should never say all **one**/`pron` knows
- *one* is a cardinal number when it could be replaced by other cardinals:
  - **one**/`numeral` of the best reasons
  - **two**/`numeral` of the best reasons

12.3 Feature `type`

12.3.1 Value `ord`

12.3.2 Value `card`

12.4 Feature `number`

Numbers other than *1* can either be singular or plural. Criteria: semantics, subject-verb-agreement.

- **9**/`sg` is similar to **6**/`sg` (referring to the symbols)
- I ordered ten dishes, but ony **9**/`pl` have arrived (referring to the entities)

12.4.1 Value `sg`

12.4.2 Value `pl`

13 I N T E R J E C T I O N S

13.1 Type hierarchy table

POS	pos	interj
		gosh
MO/SY	—	

13.2 pos-value `interj`

13.2.1 Central semantics

- Distinction for ambiguous wordforms (cf. `pos`-delimitation) is very difficult for automatic tagging. In order to minimize the spurious ambiguities, we decided to annotate only unambiguous wordforms as interjections.
- Unambiguous wordforms, i.e. such that can be nothing but interjections, are: **oh**/`interj` , **uh**/`interj` , **whow**/`interj` , **yipee**/`interj` , **gee**/`interj` , **ugh**/`interj` , **yes**/`interj` , **no**/`interj` , **gosh**/`interj` ...

13.2.2 `pos`-delimitation

Delimitation with nouns:

- **Hell**/`noun` !

Delimitation with adjectives:

- **Damned**/`adj` , I forgot my wallet
- **sure**/`adj` !
- **Great**/`adj` !

Delimitation with adverbs:

- **well**/`adv` , it was like this...
- **right!**/`adv`

Delimitation with verbs:

- **see**/`verb` , it is like this...
- if you **please**/`verb`

Delimitation with determiners:

- **my**/`det` , what a gorgeous day

14 U N I Q U E M E M B E R

14.1 Type hierarchy table

POS	pos	unique		
	type	infinitive_m	existential	negative
		to	there	not

14.2 pos-value `unique`

14.2.1 Central semantics

Single linguistic units with special distribution and/or semantics, like *not*, *to* etc.

14.2.2 `pos`-delimitation

Delimitation with conjunctions:

- in order **to**/`[conj & s_infve]` correctly specify the user needs
- I want you **to**/`infinitive_m` listen to me.

14.3 Feature `type`

14.3.1 Value `infinitive_m`

to only before an infinitive.

- It is essential **to**/`infinitive_m` not make any mistakes here.

14.3.2 Value `existential`

- Existential *there* as in
  - is **there**/`existential` enough milk left,
  - there will be rain
  - **there**/`existential` was a party **there**/`adv` .
- Impersonal *it* in stereotyped constructions:
  - weather expressions:
    it/`existential` will be raining
  - clefts **it**/`existential` was him you made you coffee.
  - extraposition **it**/`existential` is assumed that they come

14.3.3 Value `negative`

Only the word form **not**/`negative`

15 R E S I D U A L S

15.1 Type hierarchy table

POS	pos	resid			
	type	symbol	frgn_word	punctuation	unclassified
		H_2O	*per se*	*umh, er*	

15.2 pos-value `resid`

15.2.1 Central semantics

Non-linguistic phenomena, e.g. parts of another language, of graphical insertions etc.

15.2.2 pos-delimitation

Delimitation with <u>nouns</u>: foreign words are sometimes integrated (i.e. lexicalized) into the language:

- I practice **yoga**/`noun`

Delimitation with <u>conjunctions</u>:

- where they are integrated into the English sentence, they are classified as adpositions: **qua**/`adpos` , **pro**/`adpos` , **per**/`adpos` , **vis-à-vis**/`adpos` , **via**/`adpos` , **versus**/`adpos`

15.3 Feature `resid_type`

15.3.1 Value `symbol`

mathematical, scientific, technical symbols or expressions that aren't English words. BUT: names of chemicals, units of measurement (and abbreviations) are tagged as nouns!

15.3.2 Value `frgn_word`

- **mawashi**/`frgn_word`
- **Los**/`frgn_word` **Piños**/`frgn_word`
- **persona**/`frgn_word` **non**/`frgn_word` **grata**/`frgn_word`
- **bete**/`frgn_word` **noire**/`frgn_word`
- He was born **A.D.**/`frgn_word` 1978
- These artefacts date back to 50000 **B.C.**/`frgn_word`

15.3.3 Value `punctuation`

comma, ful stop, brackets, quotation marks, separating mark.

15.3.4 Value `unclassified`

For speech noises like **er**/`unclassified` , **um**/`unclassified`

References

[Atwell et al. 94] Eric Atwell, John Hughes, Clive Souter: *AMALGAM: Automatic Mapping Among Lexico-Grammatical Annotation Models*, Proceedings of SIGDAT-95, EACL-95, Dublin, 1995.

[Christ 94] Oliver Christ: A modular and flexible architecture for an integrated corpus query system. In: *Proceedings of COMPLEX'94* (3rd Conference on Computational Lexicography and Text Research). Budapest, Hungary, Jul. 1994.

[Garside, Leech, Sampson 87] Roger Garside, Geoffrey Leech, Geoffrey Sampson (eds.): *The Computational Analysis of English – A Corpus-based Approach*, Longman, London, 1987.

[Greenbaum, Quirk 90] Sidney Greenbaum, Randolph Quirk: *A Student's Grammar of the English Language*, Longman, London, 1990.

[Leech, Wilson 93] Geoffrey Leech, Andrew Wilson: *Morphosyntactic Corpus Annotation*, EAGLES, Text Corpora Working Group, Subtask 3.1: Invitation draft. University of Lancester, Dec. 1993.

[Mamrak O'Connell 92] S.A. Mamrak, C. S. O'Connell: *Technical Documentation for The Integrated Chameleon Architecture*, Ohio State University, Columbus, Mar. 1992.

[Marcus et al. 92] Mitch Marcus, Beatrice Santorini, Mary Ann Marcinkiewicz: Building a large natural language corpus of English: The Penn Treebank. – *Computational Linguistics 19,* 313–330, 1993.

[Monachini, Calzolari 94] Monica Monachini, Nicoletta Calzolari: *Synopsis and Comparison of Morphosyntactic Phenomena Encoded in Lexicon and Corpora*, Internal Document, EAGLES Lexicon Group, ILC, Università Pisa, Oct. 1994.

[Sampson 95] Geoffrey Sampson: English for the Computer. Clarendon Press, Oxford. 1995.

[Santorini 91] Beatrice Santorini: *Part-of-Speech Tagging Guidelines for the Penn Treebank Project.* Technical Report. Department of Computer and Information Science, University of Pennsylvania, Mar. 1991.

[Schulze 94] Bruno Maximilian Schulze: *Entwurf und Implementierung eines Anfragesystems für Textcorpora*, Diplomarbeit Nr. 1059, IMS, Universität Stuttgart, Feb. 1994.

[Teufel 94] Simone Teufel: *Linguistisch motivierte Corpuserschließung: Spezifikationssprache und Anfrageinterpreter*, Diplomarbeit Nr. 1058, IMS, Universität Stuttgart, May 1994.

[Teufel 95a] Simone Teufel: *A support tool for Tagset Mapping*, Proceedings of SIGDAT-95, EACL, Dublin 1995.

[Teufel 95b] Simone Teufel: *ELM-DE: A typed German incarnation of the EAGLES-TS. Definition of Lexical Descriptions and Classification Guidelines.* EAGLES report ??, Validation Action, Subtask 1. Sept. 1995.

[TEI-AI1W2 91] Text Encoding Initiative: *List of Common Morphological Features.* – TEI-AI1W2, Working Paper, Draft Version, Chicago, Jun. 1991.