



# Applying the DCR to POS Tagging

---

**Erhard Hinrichs**  
University of Tübingen  
eh@sfs.uni-tuebingen.de

Workshop on Standards, NEERI 2009

Helsinki, 09-30-09



- Goal of ISOcat/DCR:
  - Defining widely accepted linguistic concepts
- Open standard:
  - User community involvement is encouraged.
- Official standard:
  - Decision making conforms to ISO process.

- Anyone
  1. can register with ISOcat
  2. can create data categories
  3. can create data category selections (DCSs)
  4. can share DCSs
  5. can make DCSs public
  6. can submit DCSs for standardization

**Marc Kemps-Snijders, Menzo Windhouwer, Sue Ellen Wright**  
CLARIN-NL Info dag, 1 July 2009



- Existing language resources often adhere to de-facto standards
- Example: part of speech tagging
- ISOcat/DCR contains proposals for part of speech categories
- Question: What is the relation between the categories of such de-facto standards and ISOcat/DCR?

# Penn Treebank Tag Set



CC	Coord. conjunction	RB	Adverb
CD	Cardinal number	RBR	Adverb, comparative
DT	Determiner	RBS	Adverb, superlative
EX	Existential there	RP	Particle
FW	Foreign word	SYM	Symbol
IN	Prep. / subord. conj.	TO	to
JJ	Adjective	UH	Interjection
JJR	Adjective, comparative	VB	Verb, base form
JJS	Adjective, superlative	VBD	Verb, past tense
LS	List item marker	VBG	Verb, gerund / present part.
MD	Modal	VBN	Verb, past part.
NN	Noun, singular or mass	VBP	Verb, non-3rd p., sing. pres.
NNS	Noun, plural	VBZ	Verb, 3rd p. sing. pres.
NP	Proper noun, singular	WDT	Wh-determiner
NPS	Proper noun, plural	WP	Wh-pronoun
PDT	Predeterminer	WP\$	Possessive wh-pronoun
POS	Possessive ending	WRB	Wh-adverb
PRP	Personal pronoun	,	Comma
PRP\$	Possessive pronoun	.	Sentence-final punctuation

# STTS tag set



ADJ	attributive adjective
ADJD	adverbial/predicative adjective
ADV	adverb
APPR	preposition/left circumposition
APPRART	preposition with integrated determiner
APPO	postposition
APZR	right circumposition
ART	indefinite/definite determiner
CARD	cardinal number
FM	foreign-language material
ITJ	interjection
KOUI	subordinating conjunction with "zu" and infinitive
KOUS	subordinating conjunction with sentence
KON	coordinating conjunction
KOKOM	comparative conjunction

# STTS tag set



Standards  
Workshop,  
NEERI 09  
Helsinki  
09/30/09

---

VVFIN	full finite verb
VVIMP	full imperative verb
VVINFINF	full infinitive verb
VVIZU	full infinitive with “zu”
VVPP	full past participle
VAFIN	auxilliary finite verb
VAIMP	auxilliary imperative verb
VAINFINF	auxilliary infinitive verb
VAPP	auxilliary past participle
VMFIN	modal finite verb
VMINFINF	modal infinitive
VMPP	modal past participle

---

# STTS tag set



Standards  
Workshop,  
NEERI 09  
Helsinki  
09/30/09

NN	common noun
NE	proper noun
PDS	substituting demonstrative pronoun
PDAT	attributive demonstrative pronoun
PIS	substituting indefinite pronoun
PIAT	substituting indefinite pronoun without determiner
PIDAT	substituting indefinite pronoun with determiner
PPER	non-reflexive personal pronoun
PPOSS	substituting possessive pronoun
PPOSAT	attributive possessive pronoun
PRELS	substituting relative pronoun
PRELAT	attributive relative pronoun/relative pronoun
PRF	reflexive personal pronoun
PWS	substituting interrogative pronoun
PWAT	attributive interrogative pronoun
PWAV	pronominal adverb

# STTS tag set



PAV	pronominal adverb
PTKZU	“zu” before infinitive
PTKNEG	negation particle “nicht”
PTKVZ	particle part of separable verb
PTKANT	answer particle “ja”, “nein”, “danke”, “bitte”
PTKA	particle associated with adverb or adjective
TRUNC	first member of compound
XY	non-word with special characters
\$,	Comma
\$.	sentence ending punctuation
\$(	other sentence-internal signs

- DCR POS terminology consists of classes and values:

- **nounClass**

Definition: The categorization of a noun indicating whether it names a specific object or a class of objects. (ISO12620)

Explanation: Proper nouns are capitalized in English. Common nouns are not. (Mitre; TEI(green text))

- **commonNoun**

Definition: A noun or adjective denoting a class of objects. (ISO12620)

Example: continent (Mitre; TEI(green text))

- **properNoun**

Definition: A noun or adjective denoting a single object.

(ISO12620) Example: Europe (Mitre; TEI(green text))

- **adjectiveClass**

Definition: A categorization of an adjective indicating whether it pertains to a single object or to a class of objects. (ISO12620)

Explanation: Proper adjectives are capitalized in English.

Common adjectives are not. (Mitre)

- **commonAdjective**

Definition: An adjective pertaining to a generic class of objects. (ISO12620)

Example: thoroughbred horse (Mitre)

- **properAdjective**

Definition: An adjective formed on the base of a proper noun.

(ISO12620) Example: Arabian stallion (Mitre)

# Proposed ISOcat POS tag set (author: Gil Francopoulo)

---



- **Adjective:** ordinalAdjective, participleAdjective, pastParticipleAdjective, presentParticipleAdjective, qualifierAdjective
- **Adverb:** generalAdverb, particleAdverb
- **Adposition:** postposition, **preposition:** compoundPreposition, fusedPreposition, simplePreposition
- **Article:** definiteArticle, indefiniteArticle, partitiveArticle
- **Conjunction:** coordinatingConjunction, subordinatingConjunction
- **Determiner:** demonstrativeDeterminer, exclamativeDeterminer, indefiniteDeterminer, interrogativeDeterminer, possessiveDeterminer, reflexiveDeterminer, relativeDeterminer

- **Interjection**
- **Noun:** commonNoun, diminutiveNoun, properNoun, relativeNoun, voiceNoun
- **Pronoun:** allusivePronoun, conditionalPronoun, demonstrativePronoun, emphaticPronoun, exclamativePronoun, impersonalPronoun, indefinitePronoun, interrogativePronoun, negativePronoun, possessivePronoun, reciprocalPronoun, reflexivePronoun, relativePronoun, **personalPronoun:** affixedPersonalPronoun, strongPersonalPronoun, weakPersonalPronoun

# ISOcat POS tag set

---



- **Particle:** affirmativeParticle, comparativeParticle, conditionalParticle, coordinationParticle, distinctiveParticle, futureParticle, infinitiveParticle, interrogativeParticle, modalParticle, negativeParticle, possessiveParticle, relativeParticle, superlativeParticle, unclassifiedParticle
- **Punctuation:** bullet, circumposition, closeParenthesis, colon, comma, exclamativePoint, invertedComma, openParenthesis, point, questionMark, semiColon, slash, suspensionPoints
- **Verb:** be, copula, have, mainVerb

# Possible Answers:

---



- Possible answers:
  - Categories of de-facto standards are „widely accepted linguistic categories“ and as such should be registered in the ISOcat.
  - Categories of de-facto standards should be mapped onto ISOcat categories

IMHO: We need to do both.



Thank you for your attention

CLARIN has received funding from  
the European Community's Seventh Framework Programme  
under grant agreement n° 212230