

LingvoHelpilo & EspGram

- La teknika flanko



Eckhard Bick

Suddana Universitato / GrammarSoft

eckhard.bick@mail.dk

Kadro

- VISL (Universita projekto 1996-)
 - lingvoteknologio & lingvoinstruado
 - 28 lingvoj
 - analiziloj, lerniloj, ludoj ...
- GramTrans (Komerca projekto 2006-)
 - masxintradukado
 - dinamikaj vortaroj
- Lingvohelpilo (ESF-apogita projekto 2007-)
 - korektado de esperantaj teksoj

Strukturo de la prelego

1. La analizilo/parsilo (EspGram)

a) vortaro

b) senambiguiga gramatiko (CG)

c) sintaksaj arboj

2. La Korektilo (Lingvohelpilo)

a) Interfaco

b) Erarotipoj

valento kaj semantiko

- krei leksikon

- mana leksikografio
- korpusoj
- morfologio

-ig (<vt> transitivity) *colorigi (to colour), sanigi (to cure), lumigi (to light)*
-iĝ (<vi> intransitivity) *malŝanĝi (fall ill), prezidentiĝi (become president)*
-ul¹ (<H> human, person) *krimulo (criminal), ŝaŭlo (wise man)*
-in (female), -id (offspring), ge- (couple), -ist (professional) ...
-uj (<con> container), -ej (<top> placê), -a (<food>), -il (<tool>)

% vortoj klasifikeblaj per afikso

kategorio (vortoj / tipoj)	afikso	n	marktiaj %	tipoj	markitaj %
<vt> transitiva (50,358 / 12,160)	-ig	4,659	9.3 %	1,856	15.3 %
<vi/ve> ne-transitiva (34,889 / 8,675)	-iĝ	4,252	12.2 %	1,495	17.2 %
<ve> ergativa (9,688 / 1,649)	-iĝ	1,584	16.4 %	825	50.0 %
<con> ujo (1,192 / 126)	-uj	66	5.5 %	27	21.4 %
<L...> <inst> loko (12,400 / 1,806)	-ej	1,614	13.0 %	344	19.0 %
<tool> ilo (713 / 179)	-il	345	48.4 %	99	55.3 %
<H> <Hprof> <Hfam> ... homa (19,503 / 3,017) (sen la grupo-kategorio <HH>)	-ul	1,673	8.6 %	522	17.3 %
	-in	1,311	6.7 %	237	7.9 %
	-id	27	0.1 %	8	0.3 %
	-ist	2,941	15.1 %	621	20.6 %
	-an	845	4.3 %	196	6.5 %
	-estr	294	1.5 %	55	1.8 %
	-nj, -ĉj	70 (62, 8)	0.4 %	7 (3,4)	0.2 %
	ge-, bo-	79 (74, 5)	0.4 %	0	0.0 %
-ant, -int, -ont	3,158	16.2 %	719	23.8 %	
cxio homa		10,398	53.3 %	2,365	78.4

Morfologia sekcio

CG (Constraint Grammar – "Perkonteksta Gramatiko", Karlsson 1990, 1995):

- *birdo* N S NOM @SUBJ>
- *grandaj* ADJ P NOM @N<PRED
- *flugos* V FUT @FS-<ACC

Morfologia parto (4.8%): 51 REMOVE reguloj, 21 SELECT reguloj

<interr> -- <rel>: kiu, kiam, kie ...

ĉirkaŭ kvincent dolaroj - *about* 500 dollar
ĝis 15 partoprenantoj - *up to* 15 participants

PROP <-> N (malstabila ortografio en E.)

- (a) plene tradukitaj, kun finajxo: *Danio*, *Gronlando*, *Munkeno*
- (b) fonetike adaptitaj: *Buŝ* (*Bush*), *Ĥruŝĉov'o* (*Khrushchev*)
- (c) "puraj" nomoj el latin-alfabetaj lingvoj

Claude Piron, *Clinton* <--> **Piro*, **Clinto*

REMOVE (ACC) (-1C PRP) (NOT -1 PRP-DIR OR PRP-LOC)

Sintaksa CG

- aldono det sintaksaj "hipotezoj": 644 MAP reguloj (+ 29 ADD)
- ripetaj niveloj de kontekstaj senambiguigo-reguloj:
 - 541 REMOVE reguloj
 - 212 SELECT reguloj
- implicita strukturo: dependenco-sagoj

@SUBJ> (maldekstra subjekto), @<ACC (dekstra subjekto), @P<, @ADV<L

REMOVE (@<SUBJ) (*-1C @ARG/ADV<L> BARRIER VFIN)

aldono de specialaj markoj por paratagmoj (kaj, aux), elipso kaj aligo de PRP-grupoj
pludono de markoj de cxef- al relativaj frazoj, refleksivoj, pronomoj ktp

plenaj sintaksaj arboj

- PSG (frazstruktura gramatiko), kun frazeroj (funkcioj) kiel terminaloj anstataux vortoj
 - STA:fcl = SUBJ> P (<ACC, <ADVL, <SC)*
 - X:np = >N X:n N<
- DG (dependenca gramatiko) – arkoj inter *funkcio*-filino kaj *formo*-patrino
 - @FS-N< -> (αNPHEAD) IF (L) TRANS:(<rel>) BARRIER:(PR,IMPF,<co-fin>)
 - @>A -> (ADJ,ADV,DET,NUM,PCP1,STA) IF (R) NOTHEAD=(<aquant>.*@>A)
- DG pli robusta ol PSG (Bick 2005)

En [en] PRP @ADVL> #1->5

la [la] ART @>N #2->4

tria [tria] <num-ord> ADJ S NOM @>N #3->4

grupo [grupo] N S NOM @P< #4->1

kuni ĝis [kunigxi] <mv> V IMPF @FS-STA #5->0

tiuĵ [tiu] <dem> PRON DET P NOM @<SUBJ #6->5
\$, #7->0

kiuĵ [kiu] <rel> PRON DET P NOM @SUBJ> #8->9

malaprobas [malaprobi] <mv> V PR @FS-N< #9->6

ĉion [cxio] <quant> PRON INDP S ACC @<ACC #10->9
\$, #11->0

kio [kio] <rel> PRON INDP S NOM @SUBJ> #12->13

okazis [okazi] <mv> <np-close> V IMPF @FS-N< #13->10

en [en] PRP @<ADVL #14->13

la [la] ART @>N #15->17

katolika [katolika] ADJ S NOM @>N #16->17

eklezio [eklezio] N S NOM @P< #17->14

dum [dum] PRP @<ADVL #18->13

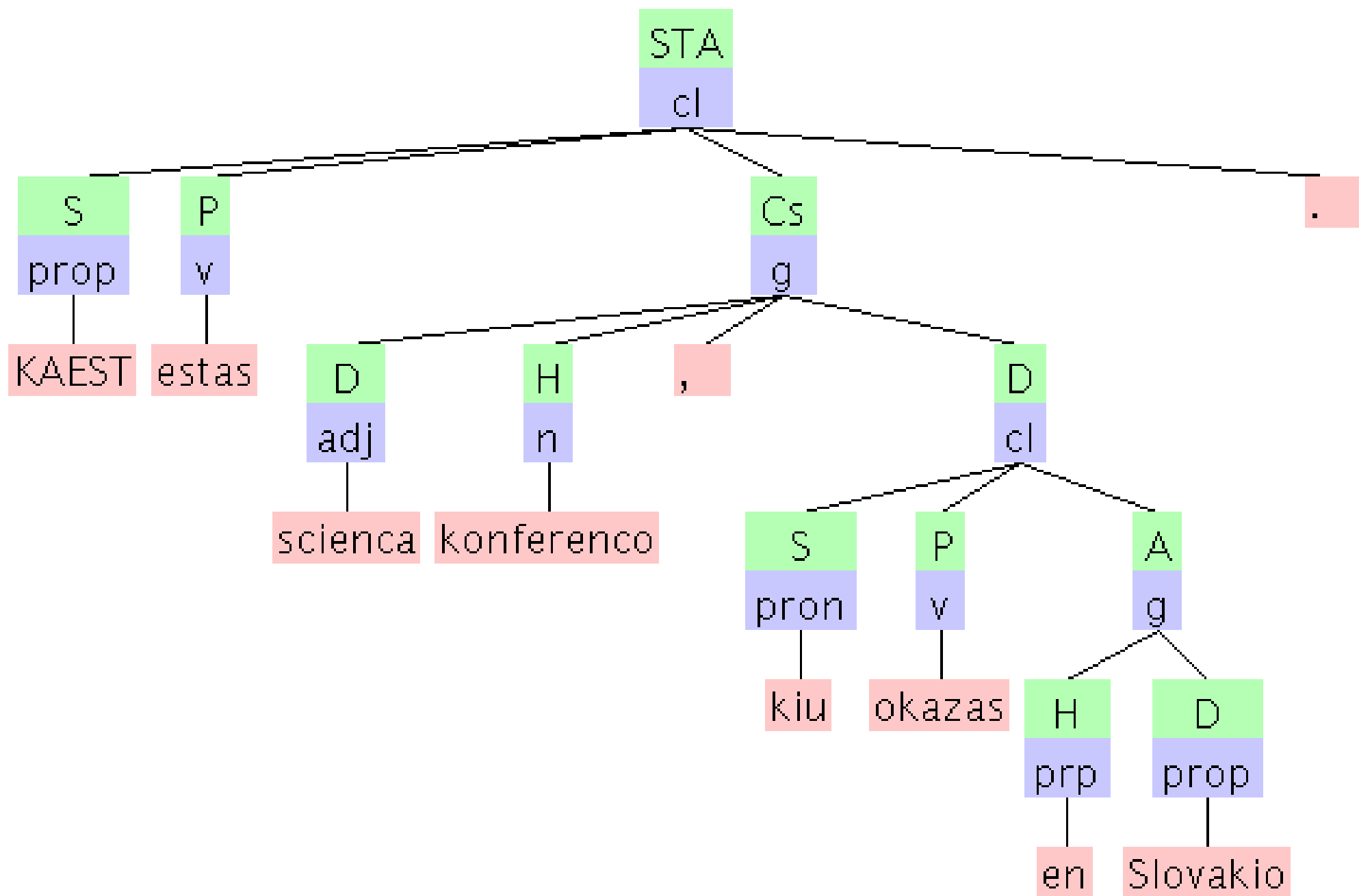
la [la] ART @>N #19->22

pasintaj [pasi] V PCP AKT IMPF ADJ P NOM @>N #20->22

dudek [dudek] <card> NUM P @>N #21->22

jaroĵ [jaro] <clb-end> N P NOM @P< #22->18

STA:fcl
=fA:pp
==H:prp("en") En
==DP:np
==DN:art("la") la
==DN:adj("tria" <num-ord> S NOM) tria
==H:n("grupo" S NOM) grupo
=P:v-fin("kunigxi" <mv> IMPF VFIN) kuni ĝis
=S:np
==H:pron-dem("tiu" <dem> DET P NOM) tiuĵ
=,
==DN:fcl
==S:pron-rel("kiu" <rel> DET P NOM) kiuĵ
==P:v-fin("malaprobi" <mv> <np-close> PR VFIN)malaprobas
==Od:np
==H:pron("cxio" <quant> INDP S ACC) ĉion
=,
==DN:fcl
==S:pron-rel("kio" <rel> INDP S NOM) kio
==P:v-fin("okazi" <mv> <np-close> IMPF VFIN) okazis
==fA:pp
==H:prp("en") en
==DP:np
==DN:art("la") la
==DN:adj("katolika" S NOM) katolika
==H:n("eklezio" S NOM) eklezio
==fA:pp
==H:prp("dum") dum
==DP:np
==DN:art("la") la
==DN:v-pcp("pasi" AKT IMPF ADJ P NOM) pasintaj
==DN:num("dudek" <card> P) dudek
==H:n("jaro" <clb-end> P NOM) jaroĵ



Kvalitoanalizo 1

(*Monato*, 4.400 unuoj, 3439 vortoj)

	Retrovo	Precizeco	F-mezuro
vortbazo / leksemo	99.7	99.7	99.7
vortklaso	99.5	99.5	99.5
morfologio, fleksio	99.7	99.7	99.7
sintaksa funkcio (@)	93.4	90.9	92.1

- komparo kun alilingvaj CG-sistemoj
 - pli bona en la morfologia kampo (kp. 1% eraroj)
 - “normala” en la sintaksa kampo (kp. cx. 5% eraroj por maturaj sistemoj, ekz.)
 - ENGCG (Lingsoft 2007)
 - DanGram (Bick 2003), Palavras (Bick 2000), HIS-Palavras (Bick 2006)
 - Estona CG (Müürisep & Uibo 2005)

Kvalitoanalizo 2

- komparo kun masxinlernaj sistemoj (CoNLL)
 - 80% (euxska) -93% (angla) korektaj funkcioj, ecx kun mane korektita morfologia analizo
- komparo kun aliaj E-sistemoj
 - Warin (2004): PDP11 kun simila morfologia korekteco (99.3%), kp. stohxastika markilo (TnT) 98.6% (ecx tiu superas rezultojn por aliaj lingvoj)
 - Lin & Sung (2004): partial parsing TBL-PSG, nekomparebla: nur frazoj de 3-5 vortoj, 1 el 30 korekta (?)

Kialoj por la sintaksa malfacileco

- libera vortordo
- internacia komunumo kun nacilingvaj interferencoj / preferoj

La? speciala komisiono de **[la]** sovetia registaro en 1943 venis al **la** alia konkludo. (*artikoloj*)

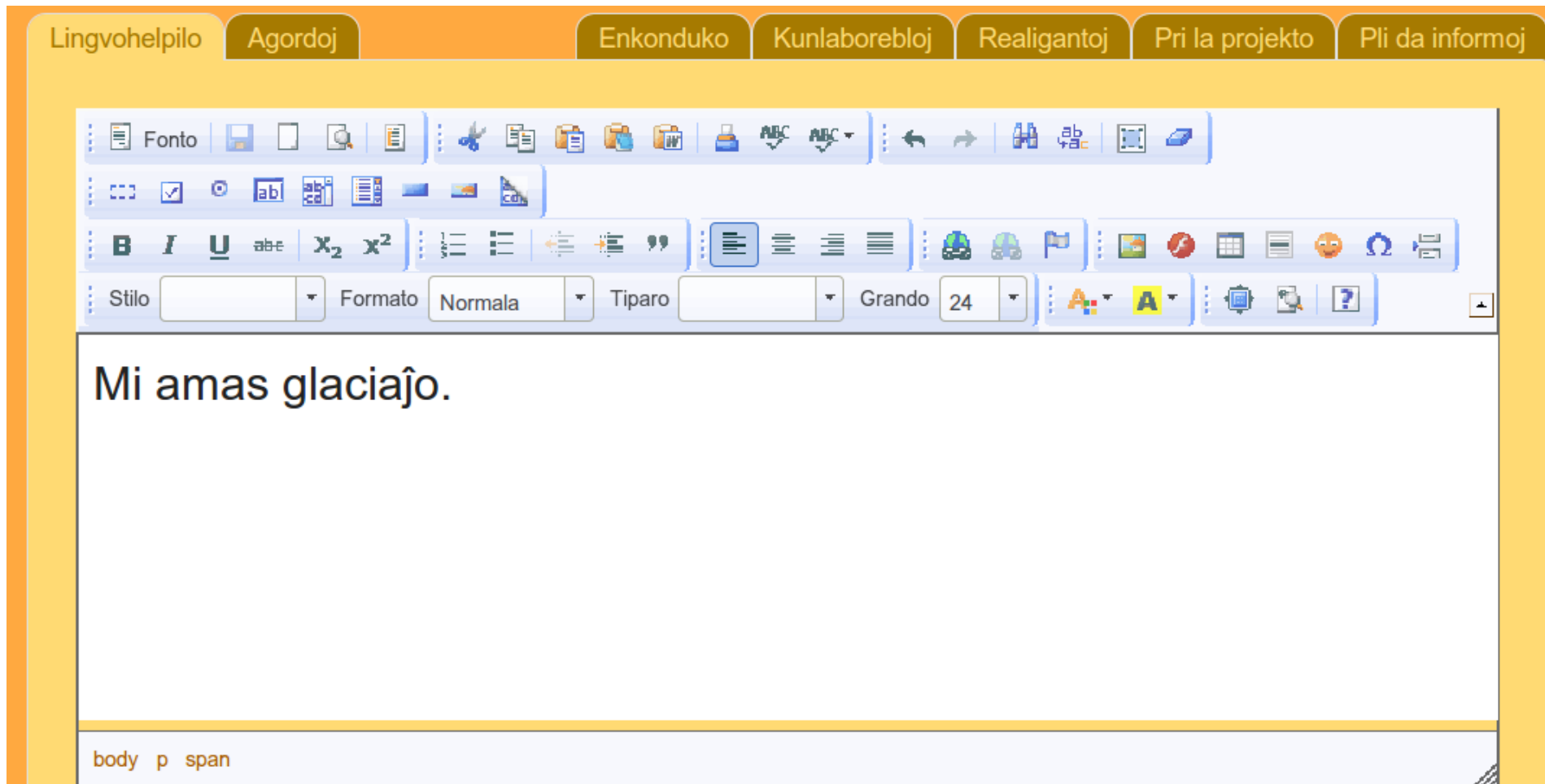
Tiel la filmo estas duoble ĝuinda de esperantistoj. (*kaj intensigo-modifo kaj frazkomplemento*)

Nun jam planite estas, eksporti la filmon al la tuta mondo. (*pre-lokado de la participo*)

Lingvohelpilo

- plurjara projekto, subteno de ESF, entute 5-6 personoj
- Celo: ortografia + gramatika korektilo por Eo, unue per-rete
 - traktado de kunmetajxoj kaj afiksado
 - traktado de ekzistantaj, sed kontekste malgxustaj formoj
 - pedagogio: eraro-kategorio kun klarigoj
 - ordigo de korektoproponoj laux frazokonteksto
- Uzo de la EspGram-analizilo por liveri gramatikan kontekston
- Uzo de CG-reguloj (vislcg3) por marki erarojn
- Uzo de spertoj el alia (dana) CG-korektilo
- Kolektado de erarekzemploj, kompilo de erarkorpuso
- Pedagogiaj komentoj laŭ erarklasifigo

Grafika edakto-interfaco



Kontroli

Raportu problemojn pri frazanalizoj al "eckhard.bick" ĉe "mail.dk"
Raportu teknikajn problemojn de la interfacoj al "bertilow" ĉe "gmail.com"

Kelkaj falsaj/mankantaj erarmarkoj okazas pro pure retumilaj/interfacaj kialoj. Plej sekure estas enmeti nur unu frazon, ne plurajn. Eventuale komparu kun la grafike pli simpla universitata interfacoj, kiu donas pli rektan aliron al la korektiloj mem (uzu x-cxapelojn).

Erarmarkado kun proponoj kaj klarigoj

Lingvohelpilo

Agordoj

Enkonduko

Kunlaborebloj

Realigantoj

Pri la projekto

Pli da informoj

Mi amas glaciaĵo.

glaciaĵon
glaciaĵo

La lingvohelpilo trovis unu lokon, kiu eble estas erara.
Eble vi volas korekti ĝin.

	Gramatikaj	Literumaj
Molaj	0	0
Malcertaj	0	0
Sume	1	0

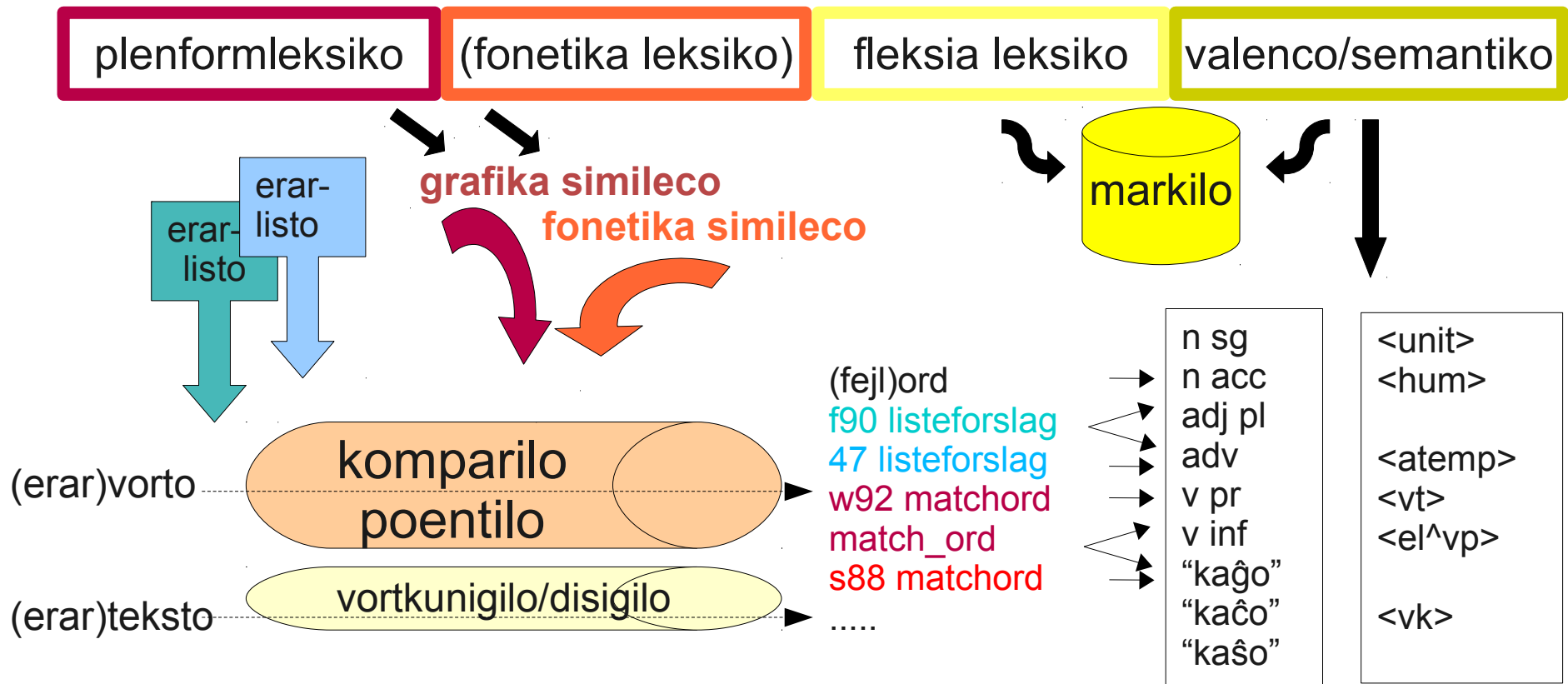
akuzativo, objekto

Li manĝis pomo.

→ **Li manĝis pomon.**

La rektan objekton oni indikas per -n.

<http://lernu.net/lernado/gramatiko/konciza/objekto.php>

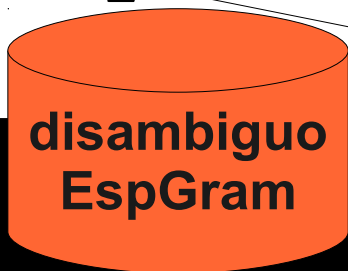


EspGram 1

CG map-reguloj

CG remove/select

fleksia leksiko



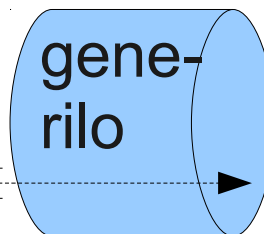
konteksto

kor-1 Analizo 1a
Analizo 1b
kor-2 Analizo 2
kor-3 Analizo 3a
Analizo 3b
kor-4 Analyse 4
kor-5 Analyse 5a
Analyse 5b

@majusklo
@comp
@headstop
@:propono

Analyse 1ab
Analyse 2
Analyse 3ab
Analyse 5ae
Analyse 5b
Analyse 4

@vfin
@inf
@pl
@acc



**Ret-5b
fleks.**

1. propono
2. propono
3. propono
4. propono
.....

► komparilo-ordigita restpropono

EspGram 2

Grammar checker examples

- ADD (%nom) TARGET ACC + @ACC> (s @<ACC) (NOT s @SUBJ) ; # fleksia eraro en subjekto: **Lin** faris kukon
- ADD (%pl) TARGET ("kiu" S) (p V LINK pS @FS-N< LINK p P) ; # agordo por relativaj pronomoj: aperas **personoj** tie kaj cxi tie **kiu ekkonis**
- ADD (%acc) TARGET (N NOM) (*-1 >>> BARRIER @NON->N) (*1 @SUBJ> BARRIER @NON->N OR KC OR KOMMA) (p (<vt>) LINK NOT c @ACC OR @SC) ; # frazkomenca objekto rekonita pere de verbo-transitiveco, manko de objekto kaj ĉeesto de subjekto: **Profundo** vi povas altigi ...

ortografio

marko	signifo	ekzemplo
: ... (vorto)	propono de nova vortformo: (a) ortografia eraro (b) malgxusta vortelekto	(a) espernatisto [esperantisto] (b) sur [en] mia retejo kostas 8 dolaroj po [por] metro (ADV->PRP) sxi estas amata far [de] cxuij

Fleksivoj 1

marko	signifo	ekzemplo
pl, pl* ¹	pluralo	abaux virino [virinoj]
sg, sg*	singularo	li estas junaj [juna]
acc	akuzativo	li mangxis pomo [pomon] du jaroj [jarojn] antaux la milito sxi preferas la francan kiel dua lingvo [lingvon]
nom	nominativo	cxeeestas tie multajn naciajn tradukojn [al] mi estis donita trinkajxon [trinkajxo] pluvis sxtonojn [sxtonoj] sur ilin
nom + al	“dativo”	Mi dankis lin [al li] la tujan reagon.

-
- 1 cxe nombro-eraro en nominalfrazo, normale la reguloj komparas modifikatorojn (ADJ, DET) kaj kemojn (N), kaj plej ofte ne eblas certe diri kiu el la du estas erara, do ambaux estas markitaj. Tamen, pro la semantika pezo de la kerno, estas malpli versxajne ke estas gxi kiu adaptigxu al la modifikatoro. En tiu kazo, ni uzas asteriskon post la @sg aux @pl markoj.

Fleksivoj 2

marko	signifo	ekzemplo
inf->vfin	finitivo (-as, -os, -is, -us) from infinitive	kredas ke religio esti [estis] la kreado de Dio Li rigardis televidon dum mangxi [dum li mangxis] Li ne povis skribi pro ne havi inkon [cxar li ne havis]
abs->vfin	finitivo (-as, -os, -is, -us) from participle small clause	La infanoj veninte hejmen, li povis trankviligxi (post kiam la infanoj venis hejmen) Farite/farita la laboro, ili iris hejmen (Post kiam ili faris la laboron, ...)
is	pasinteco (-is)	antaux 50 jaroj en Euxropo preskaux ne ekzistas [ekzistis] ... li jxus venas [venis]
us	imago-formo (-us)	skribunta viro, skributa letero, li estis skribunta (li estus skribinta)

Fleksivoj 3

marko	signifo	ekzemplo
inf	infinitivo (-i)	la politikistoj tion volas sxangxos [sxangi]
cond	ne-realo (-us)	se mi estu [estus] ricxa, mi estus felicxa
ita	perfektiva aspekto	Sxi estis naskata [naskita] la kvaran de januaro
ata	imperfektiva aspekto	Sxi estis plu kaj plu amita [amata] La foiro estas cxiam vizitita [vizitata] de multaj homoj.
inte	adverba participo	Post diri tion [Dirinte tion]
akt	aktiva (ne pasiva) formo	Paroli estis finite de li [Li finis paroli]

Vortklasoj

marko	signifo	ekzemplo
iu	tabelvort-eraro	cxio [cxiu] granda religio
adv	adverbo	Estas normala [normale] ke ... Du jaroj antaux [antauxe] mi estis en Islando
adj	adjektivo	Tio estas bone [bona]. pripensi viroj kaj inoj interrilatoj
noun	substantivo	... por kiu cele [celo] ni pagos
ordinal	ordinala numeralo	la 20 [20-a] jarcento
refl	mankanta refleksivo	li preferas lian [sian] propran biciklon
no-refl	troa refleksivo	sia [lia] edzino estis kisata de li Sxi sentis, ke pluvas sur sin [sxin]

Afiksoj

marko	signifo	ekzemplo
ado	-ado finajxo	du monatoj de studi [studado] de Esperanto
ig	-ig transitiveco	mallaulegxante [...legxigante] alkoholon lumita [lumigita]
ul	-ul substantivigo	li estas genio [geniulo]
DEL:ul	duobla +HUM	skribantulo [skribanto] virulo [viro]
culture	nacikulturaj influoj	socialdefendo [socia monsubteno?]

“preter-lingva” ortografio

marko	signifo	ekzemplo
lower	minusklo	en Hungara [hungara] popolarmeo
upper	majusklo	en italio [Italio]
comma	mankanta komo maldekstre	Sxi rakontis [,] ke sxi estis en Afriko.
no-hyphen	tro ligstreko	cxi-tie (cxi tie)
E-TYPE:dis	erara kunskribo	pied pilko [piedpilko] <i>markita</i> : pied=pilko
E-TYPE:joined	erara dis-skribo	novaloko [nova loko]

en- kaj forigo de vortoj

marko	signifo	ekzemplo
remove	troa vorto	en la mia domo [en mia domo] estas kostus multe da mono estas normale voki geamikojn kiel gefratoj la urbo de Novjorko [la urbo Novjorko]
PREADD:.. (vorto)	mankanta vorto maldekstre	Vi volas veni? [Cxu vi volas veni?] interkonatigxi multajn amikojn [kun multaj amikoj] mi ne parolas cxinan [la cxinan] Mi dankis lin la tujan reagon [al li]
PREADD:subject	mankanta subjekto	Mi diris, ke venu al mi [ke vi/li ... venu al mi] Sxi rakontis, ke [sxi] estis en Euxropo.
@:... (vorto) @:BASE-... (bazo)	malgxusta vortelekto (BASE transprenas la fleksivojn de la originalo)	pri kiu subjekto vi diskutis? [temo] Lia heroeco tre sxatis al mi [tre placxis al mi]

sintakso

marko	signifo	ekzemplo
clefting	emfazo-strukturo uzata ekster en skandinavaj lingvoj	Tiu laboro mi pensas, ke estas malfacila [n pensas, ke tiu laboro estas malfacila]

“molaj” eraroj

marko	signifo	ekzemplo
<softerror>	normala uzo cxe kelkaj esperantistoj	skributa letero tiom forta kiom vi iomete da teo amata far cxiuj

Reto:

<http://beta.visl.sdu.dk>
(EspGram & Lingvilo)

<http://corp.hum.sdu.dk> (CorpusEye)

<http://gramtrans.com> (DeepDict)