

Esperantilo

<http://www.esperantilo.org>

Artur Trzewik

Gramatika Kontrolilo Kunsido
Stokholmo 08.2007

Agendo

- Historio de projekto Esperantilo
- Libera programado
- Trajtoj kaj uzado de la programo
- Sintaksa Analizo
- Gramatika Korektado
- Tekniko de programo
- Aktualaj temoj kaj interesoj
- Spertoj dum programado de gramatika korektilo

Historio

- Projekto naskiĝis en 2003 kiel simpla UTF-8 redaktilo kun x-codo transformo. Unue projekto nomiĝis EsperantoEdit.
- Facila instalado. Ĝi kuras sur Linukso kaj Vindozo.
- Samtempa lernado de Esperanto kaj komputila lingvistiko. Malrapida envolvo de programo

Libera programado

- Movado de programistoj.
- "Free Software" signifas "Libera" ne "senkosta" programaro.
- GNU-Programoj, Richard Stallman
- Ekzemploj: GCC compiler, Linukso, Mozilla (FireFox), OpenOffice, Apache, Wikipedia, REVO-Vortaro
- Programistoj estas ofte Uzantoj. Uzantoj estas ofte Programistoj.
- Komunumo de Programistoj kaj Uzantoj

Leĝa bazo GPL-Permesilo

- Rajto je fonta kodo
- Rajto je ŝanĝo kaj plua evoluo
- Derivata programaro devus havi la saman permesilon (kontraŭe al BSD-Permesilo)
- Neniu garantio je funkciado
- Neniu garantio je kromaj efektoj

- Simileco al Permesilo de d-ro Zamenhof rilate al Fundamento! (Scio estas propraĵo de ĉiuj)

Trajtoj de Esperantilo

- Redaktilo
- Literumado
- Vortaroj
- Traduka Sistemo (Traduka memoro)
- Gramatika Korektilo
- Lingvistikaj iloj

Redaktilo

- UTF-Redaktilo
- Konvertilo inter diversaj kodoj (Unicode, iso-8559-2, KOI8-R)
- Subteno de formatoj: HTML kaj XML
- Viki-Formato
- Mutaj esperantaj provizoraj skribmanieroj: gx, gh, g[^], [^]g, HTML.

Literumado

- Speciala aliro. Rekono de esperanta vortfarado. Du ŝtupoj de eraroj. (Eblaj sed ne konataj vortoj)
- Sugestoj surbaze de programo kaj vortaro, kiu estas uzata en OpenOffice (Hunspell)
- Vortaro baziĝas sur provizo de radikoj kun prefiksoj kaj sufiksoj.
- Redaktilo de radikaro.

Vortaroj

- Multlingva vortaro. Esperanto estas la baza lingvo. Eblo de Redakto (Importado, Eksportado)
- Aldonaj priskriboj pri semantiko, gramatiko kaj fleksio. (Por unuopaj lingvoj)
- Montrilo de REVO-Vortaro
- Vortaro de sinonimoj (Tezaŭro)

Traduka Sistemo

- Funkcias bone nur por pola lingvo (limigitaj ebloj por germana kaj angla lingvoj)
- Tradukado sur nivelo de sintakso
- Traduka memoro
- Subtenas diversaj formatoj: HTML, OpenOffice, Word, XML, XLIFF
- Interaktiva tradukado

Gramatika Korektilo

- Korektado de tuta teksto
- Korektado po unu frazo
- Sugestoj de korekto
- Korektado surbaze de sintaksa analizo
- Korektado surbaze de ekzemplaj esprimoj (pattern matching)

Lingistikaj iloj

- *Tokenizer*
- *Tagger*
- *Syntax Parser*
- Statistikaj iloj (Frekventeco)
- Bazo de frazpartoj

Gramatika korektilo

Ekzistas du manieroj de korektado

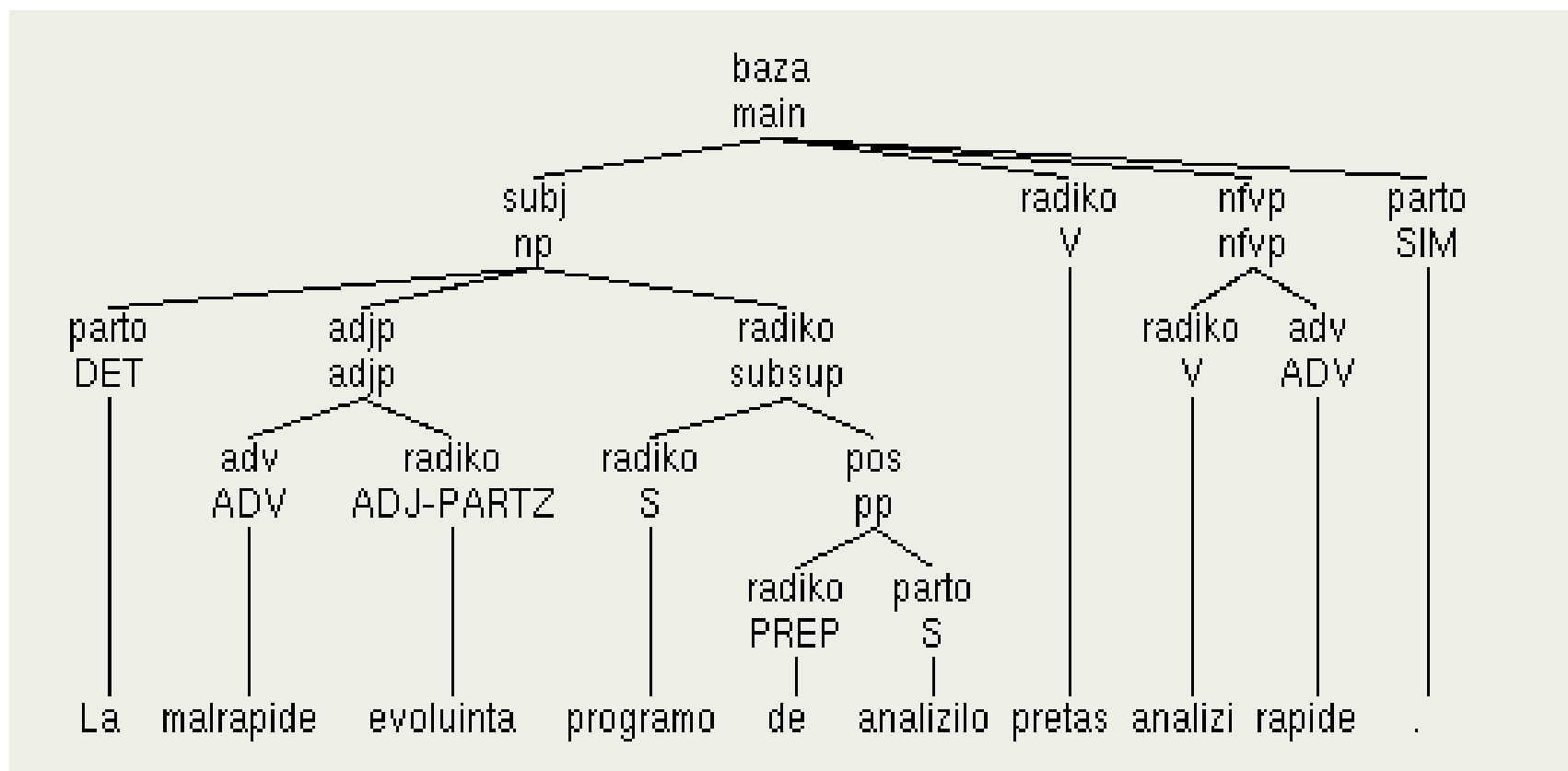
- analizo de sintaksa arbo
- analizo laŭ ekzemplaj esprimoj (Pattern matching)

Sintaksa Analizo

- "Bottom-Up" algoritmo
- Sen "backtracking"
- Analizo estas operacioj sur arboj
- Uzado de objektema programado
- Elementoj de arboj estas objektoj (objektema programado). Reguloj de analizo estas enprogramita en tiuj objektoj
- Parta analizo de malkorektaj frazoj estas ebla

Programado de sintaksa analizilo

- 118 ekzemplaj reguloj
- Aldona programado en programa lingvo



Defioj de sintaksa analizo

- Multifunkciaj vortetoj: (dum, kiel, al, de)
- Idiotaĵoj (unu post la alia, eble, de alte, ...)
- Infinitivaj frazpartoj kaj participoj
- Struktura Ambivalenco
 - Ŝi ricevis la leteron de amiko.
letero de amiko aŭ amiko donis la floron.
 - Li promesis al ŝi veni.
veno al ŝi aŭ promeso al ŝi.

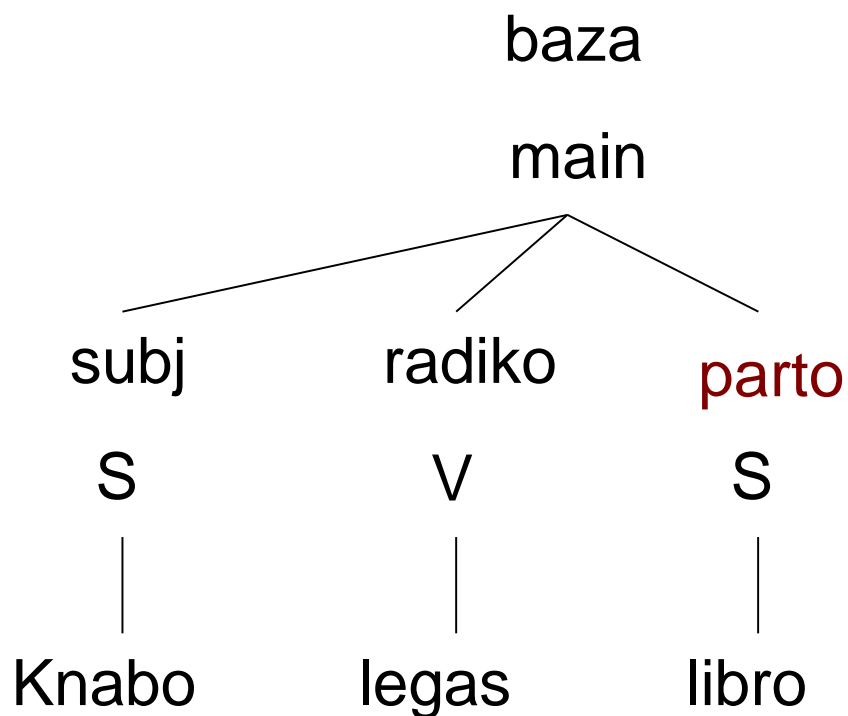
Analizo de sintaksaj arboj

Troveblaj eraroj

- akuzativo
- harmonio de nombro kaj kazo
- misuzo de verboj ("valency thoery")
- Interpunkcio

La korektado estas programita kiel "visitor pattern".

Ekzemplo 1 - akuzativo

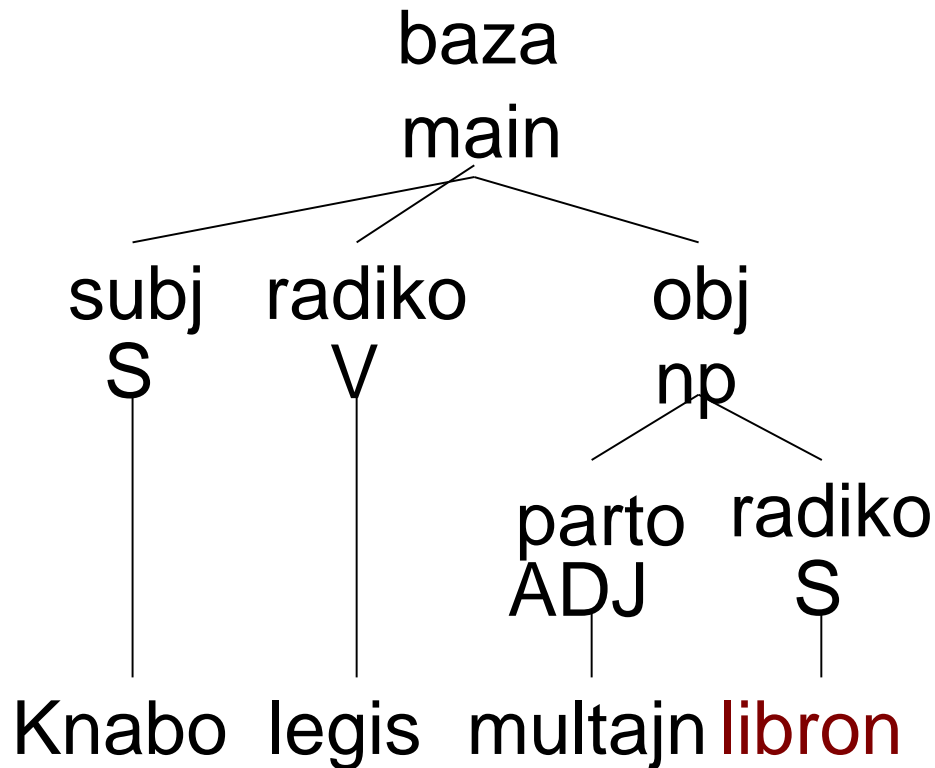


Elemento "libro" ne havas funkcion en la frazo.
Verbo "legi" estas transitiva. Objekto ne ekzistas.
Do "libro" povus esti erara akuzativo

Defioj de akuzativo - Ekzemploj

- Mi laboras libron. (labori ne estas transitiva)
- Mi laboras tutan tagon.
- Li iras sian vojon. (Li iras la kapon)
- Oni elektis lin prezidanto.
- Vidante ŝin li ridetis.
- Li venis por vidi vin.
- Li venis al domon.
- Li estas instruiston.
- Li havas komputilon kaj libro.

Ekzemplo 2 – harmonio de nombro



Branĉo "np" testas la nombron de elementoj.
Elementoj "multajn" kaj "libron" ne harmonias.

Ekonomio de korektado

- Teorie la nombro de eraraj frazoj estas pli granda ol nombro de korektaj frazoj.
- Statistiko helpas ĉe elekto de la plej oftaj eraroj.
- Esperantistoj devenas de diferencaj kulturoj, do eraroj ankaŭ diferencas.
- Statistika materialo estas necesa.

Resumo

- Kvalito dependas de bona sintaksa analizo
- Ofte falsaj asertoj pri eraroj
- Ĉiu tipo de eraro devas esti programita speciale
- Genero de bonaj sugestoj estas malfacila

Ekzemplaj esprimoj

Troveblaj eraroj

- Kutimaj eraroj (pro timi)
- Evitindaj vortoj (teĥniko)
- Specialaj kazoj (dekunu)
- Vortuzado (interesi, ...)

La sama ideo estas uzata en programo
"LanguageTool" por OpenOffice

Ekzemploj de Esprimoj

- Nomoj de dosieroj
*.txt esp-?.xml
- "Regular expressions" ofte ĉe tekstaraĵoj
.+oj?n? - ĉiuj substantivoj
parol(ad)?o – parolado aŭ parolo

Esprimoj en Esperantilo

- Elementoj estas vortoj ne signoj
- La informoj de etikietilo (tagger) estas alireblaj
- Propra difinado

Ekzemplo

```
folioj {  
    v pro  
    s V  
}
```

Tio signifas: Du vortoj. La unua vorto estas "pro" la dua vorto estas verbo.

Eraroj: "Mi ne venas pro timi".

v – vorto, s – simbolo, b – baza vorto

Logikaj operacioj estas eblaj (AND OR NOT)

Ekemplo – kompleta regulo

```
regulo {  
  nomo {pro timi}  
  match {  
    folioj {  
      v pro  
      s V  
    }  
  }  
  mesagxo {post "pro" ne povas esti infinitivo}  
  regiono 1  
  pkorekto {[string range $v2 0 end-1]o}  
}
```

Tekniko de programo

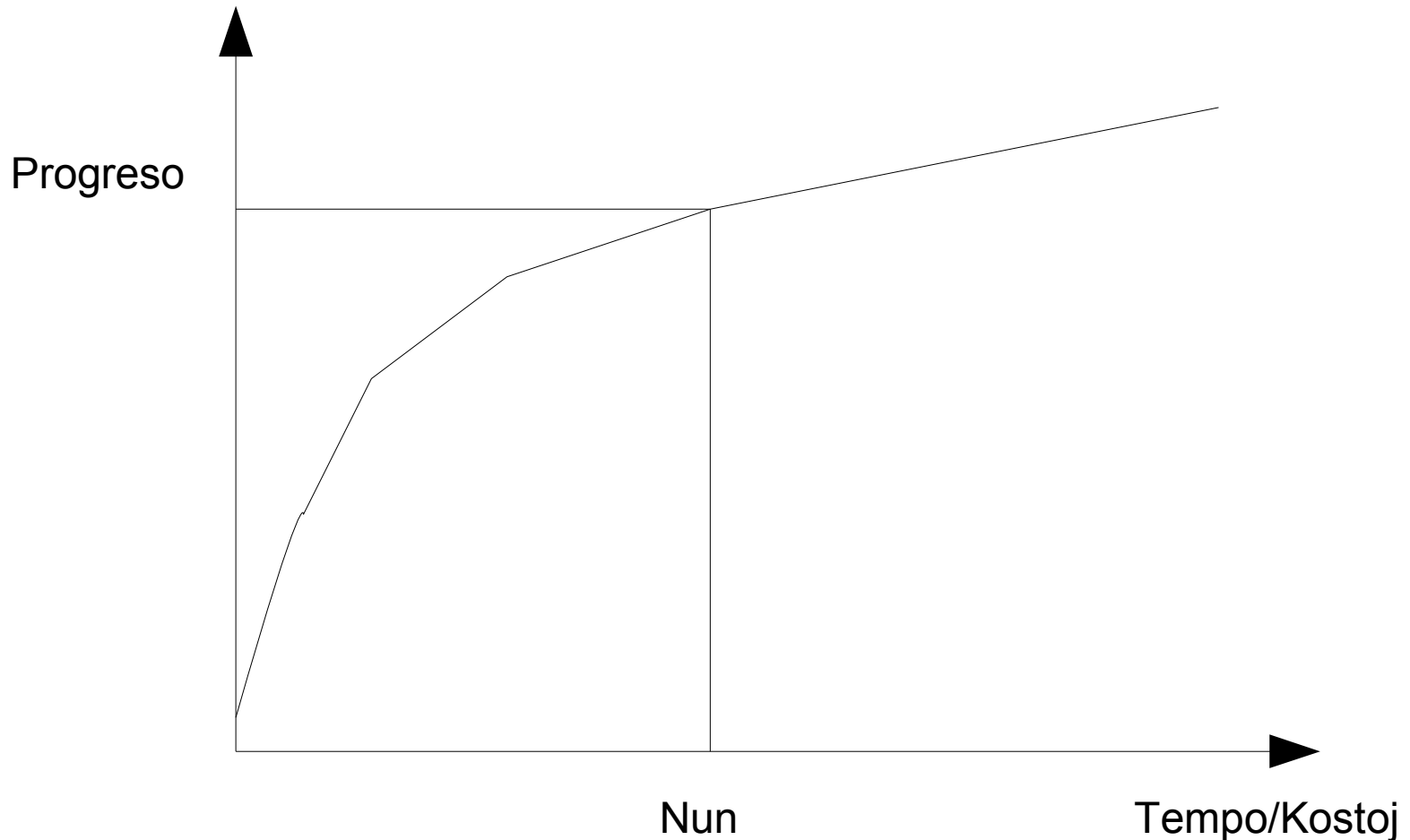
- Uzas objekteman programlingvon de sekva generacio (XOTcl)
- XOTcl baziĝas sur tre facila Tcl
- Interaktiva programa medio kun vastaj ebloj de inspektado de kuranta programo.
- Tre dinamika programa lingvo sen tipoj.
- Ĉiuj teknikaĵoj estas liberaj kaj senkostaj
- Bone laboras kun C/C++.
- Multaj aldonaj moduloj ekzistas (XML, GUI, ZIP, SQL, Hunspell)

Tekniko de programo 2

- La tekniko ne estas vaste konata
- La tekniko ne estas simila al plej konataj nunaj sistemoj (Java, C#)
- Eble neniam uzebla direkte en profesia laboro.

Spertoj post 4 jaroj de programado

- Progreso estas ankoraŭ ebla sed kostas pli kaj pli



Uzantoj de Esperantilo

- 50 elŝutoj pro semajno
- Problemoj pri Kunlaboro (Kontakto)
 - Spertaj Esperantistoj ne bezonas tian programon
 - Komencantoj ne kapablas helpi
 - Libera programado ne estas vaste konata
- Multaj uzantoj havas tre malgrandajn spertojn pri komputiloj kaj programaro.
- Multaj uzantoj estas altaĝaj (pensiuloj)
- Tre altaj atendoj al programo. Lingvo estas tipa domeno de homo (Programo ne estas ja perfekta)

Uzantoj de Esperantilo (2)

- Por multaj homoj la programo estas la unua kontakto kun Esperanto (Lingvistoj, Programistoj)
- "Esperantilo" estas menciita en 2 grandaj komputilaj ĵurnaloj (Linux Magazine)
- Funkciado de Programo povus influi la evoluon de lingvo (Komencantoj ofte tre fidis al programo)

Aktualaj temoj por eldono 1.0

- Maŝina tradukado el pola lingvo
- Korekto de eraroj (raportitaj eraroj!)
- Plibonigo de uzula interfaco kaj uzebleco
- Plibonigo de maŝina tradukado

Eksperimentoj kaj interesoj

- Semantikaj retoj (kiel konata WordNet)
- Kreado de paralelaj korpusoj
- Semantika korektado
- Kunlaboro kun OpenOffice, FireFox kaj KDE
- Subteno de Mac OS
- Direkta konekto al interreto (vortaroj)

Lingvaj problemoj dum Programado

- Mankas vastaj liberaj lingvistikaj materialoj (vortaroj, tekstaroj, glosaroj)
- Esperantologio estas juna disciplino
- Esperantistoj estas ofte veraj lingvaj fanatikuloj. Tial pragmatika helpo je iu problemo ofte ne estas ebla. Opinioj pri korekteco kaj bona esperanto estas parte tre diferencaj.

N. Chomsky kaj Fundamento

- Ĉu Esperanto havas generativan gramatikon?
- Respondo: "*Esperanto ne estas lingvo, sed vortaro kaj sugesto pri gramatiko. Homoj komprenas ĝin ĉar ili uzas siajn proprajn intuiciajn sciojn. Do, al la demando ne eblas respondi,*"

Fundamento ne priskribas la generativan gramatikon. Kelkaj esperantistoj asertas, ke gramatiko estas jam denaske difinita.

Korekteco de Frazoj

- Esploroj de Hiroshi Nagad kaj R. Corsetti (Influoj de gepatra lingvo sur la lernadon de esperanto: psikolingvistika esploro)

Korekteco = Kutimeco

Mi havas libron bonan.

Rapidan mi havas hundon.

Mi la fraton volis viziti.

Rezulto de esploroj: La sento de korekteco dependas de nacia lingvo kaj sperto pri Esperanto.

Aldonaj reguloj

- Li parolis pri ĉirkaŭ 10 homoj.
- Kiel vi fartas?
- Mi laboras kiel vi.
- Oni venis unu post la alia.
- Li estas ema fari tion.
- Estas eble fari tion.
- Li parolis pri eble 3 libroj.

Naciaj kutimoj

- *Li vidis lin paroli.*
- *Li komencis tralegi la libron. (Eraro?)*
- *Li helpis vin. Li helpis al vi.*
- *Ili ne konis unu alia.*
- *La birdo falis de alte.*
- *Li atingis la sian.*

Ekzemploj - Azianoj

... *pro mia ne ankoraŭ sufiĉe bona kono ...*

Li jam ne laboras.

Por Germanoj

Li ne povis fumi = Li devas ne fumi

Sintaksa Analizo de Esperanto

- Mankos bankoj de sintaksaj arboj
- Nur PMEG enhavas grandan kolekton de frazoj. (sed ne en formato por komutiloj) – PMEG ne enhavas sugestojn pri bona (korekta) uzado
- Ne ekzistas formala priskribo (kiel Lojĵbano)

Interpunkcio en Esperanto

Definitaj reguloj ne ekzistas. Naciaj kutimoj estas uzataj.
Ekzemple "angla" (mis)uzado de komoj.

- *Mi scias ke li venis.*
- *Li vidis libron pri kiu li parolis tre ofte.*
- *Pro manko de mono, li ne venis.*

En angla lingvo pro rigida ordo de vortoj komoj ne estas tiel gravaj kiel en Esperanto.

Fino

Diskuto kaj Demandoj?