

# Kiel komputilo tradukas?

Klarigo per ekzemplo tradukata al pola lingvo

Artur Trzewik

[www.esperantilo.org](http://www.esperantilo.org)

# Celo de Prezento

*itala proverbo:* Traduttore, traditore  
Tradukanto, perfidanto

- rapide enkonduki en maŝina tradukado
- montri la teknikon uzatan en programo "Esperantilo"
- plivastigi la komprenon por problemoj kaj defioj de maŝina tradukado

# La Fonto

„La maljuna virino legis bonan libron hieraŭ.“

Por komputilo teksto estas ĉeno de nombroj. Ĉiu nombro signifas unu literon

76, 97, 32, 109, 97, 108, 106, 117, 110, 97, 32, 118,  
105, 114, 105, 110, 111, 32, 108, 101, 103, 105, 115,  
32, 98, 111, 110, 97, 110, 32, 108, 105, 98, 114, 111,  
110, 32, 104, 105, 101, 114, 97, 365, 46

La rilato inter literoj kaj nombroj oni nomas „enkodo“. Por esperanto oni uzas kutime tiel nomatan UTF-8 enkodon.

# "Tokenize" – Serĉu vortojn



La maljuna virino legis bonan libron hieraŭ.

Vortoj

La maljuna virino legis bonan libron hieraŭ .

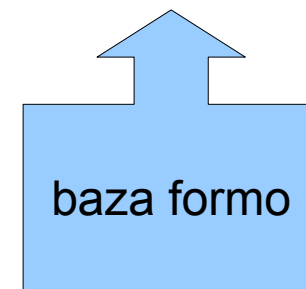
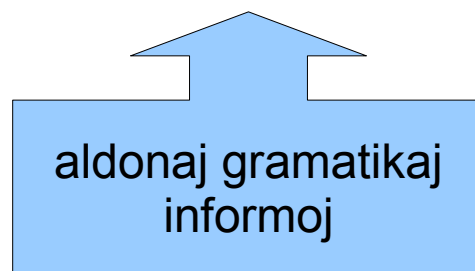
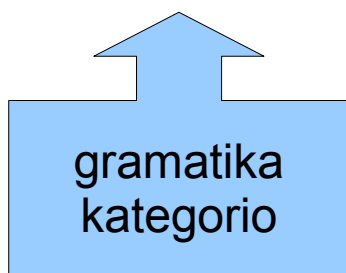
Frazo

La tasko de „Tokenize“ estas trovi vortojn kaj frazojn.

# Etiketado (*Tagging*)

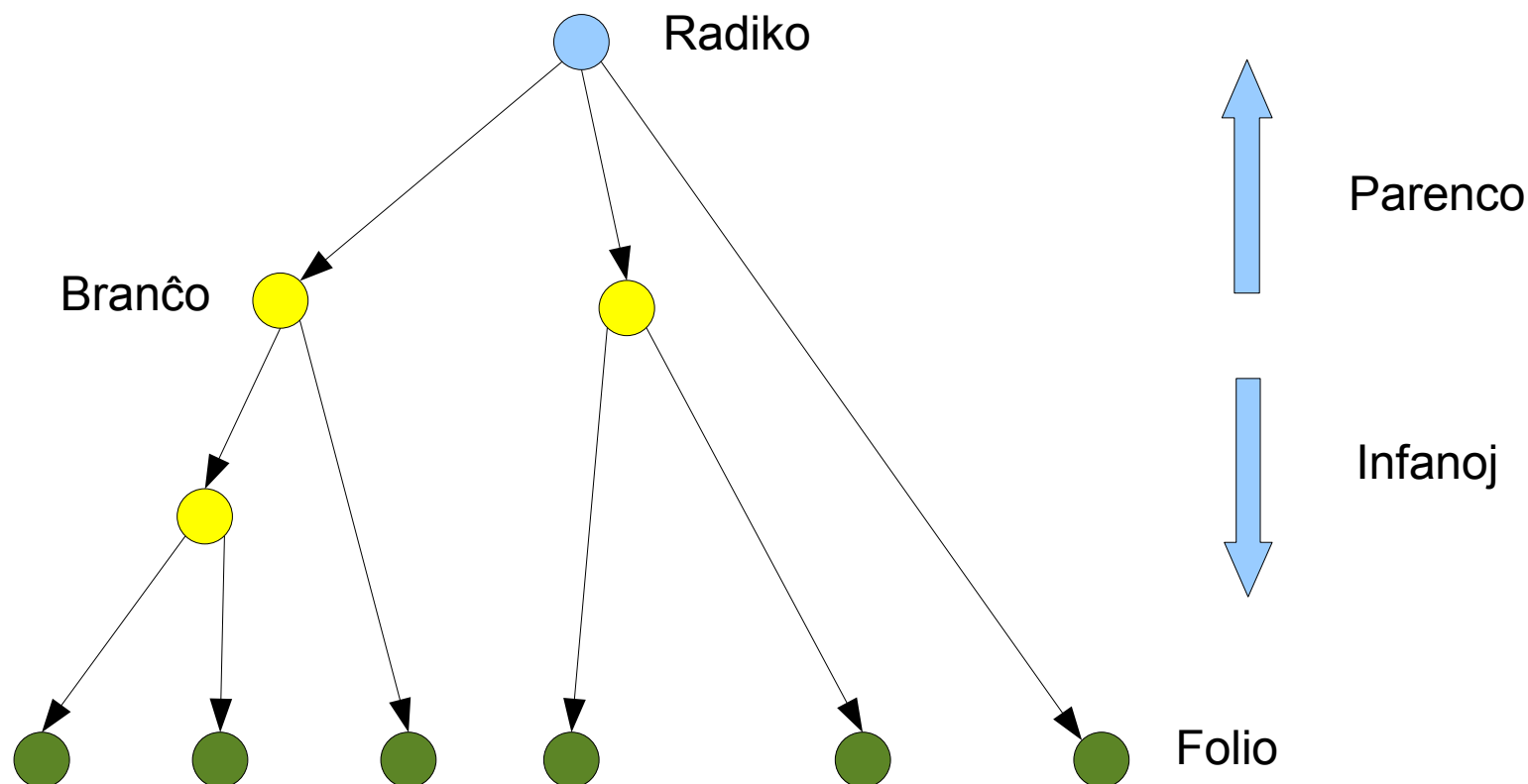
En tiu ŝtupo vortoj estas analizitaj po unu laŭ gramatika deveno

La	determinilo		la
maljuna	adjektivo	kazo=nominativo, nombro=singularo	maljuna
virino	substantivo	kazo=nominativo, nombro=singularo	virino
legis	verbo	tempo=is	legi
bonan	adjektivo	kazo=nominativo, nombro=singularo	bona
libron	substantivo	kozy=nominativo, nombro=singularo	libro
hieraŭ	adverbo		hieraŭ
.	simbolo		.



# Sintaksa Analizo - Enkonduko

Kutime la strukturo de frazo estas prezentata kiel sintaksa arbo.  
Arboj estas en matematiko specialaj direktitaj grafeoj.  
Oni diferencas: radikojn, branĉojn kaj foliojn kiel elementoj de arbo.

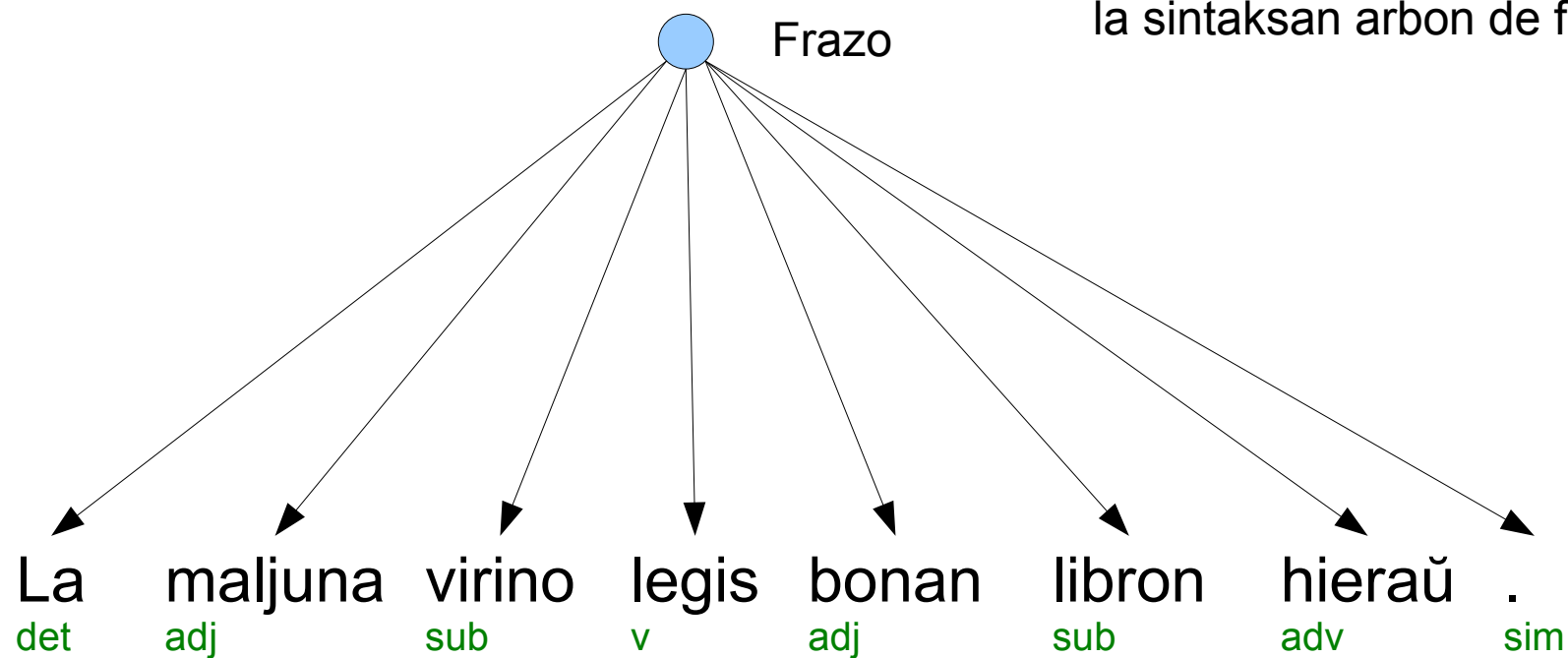


# Sintaksa Analizo – Sintaksa Arbo

En sintaksa analizo de tekstojn:

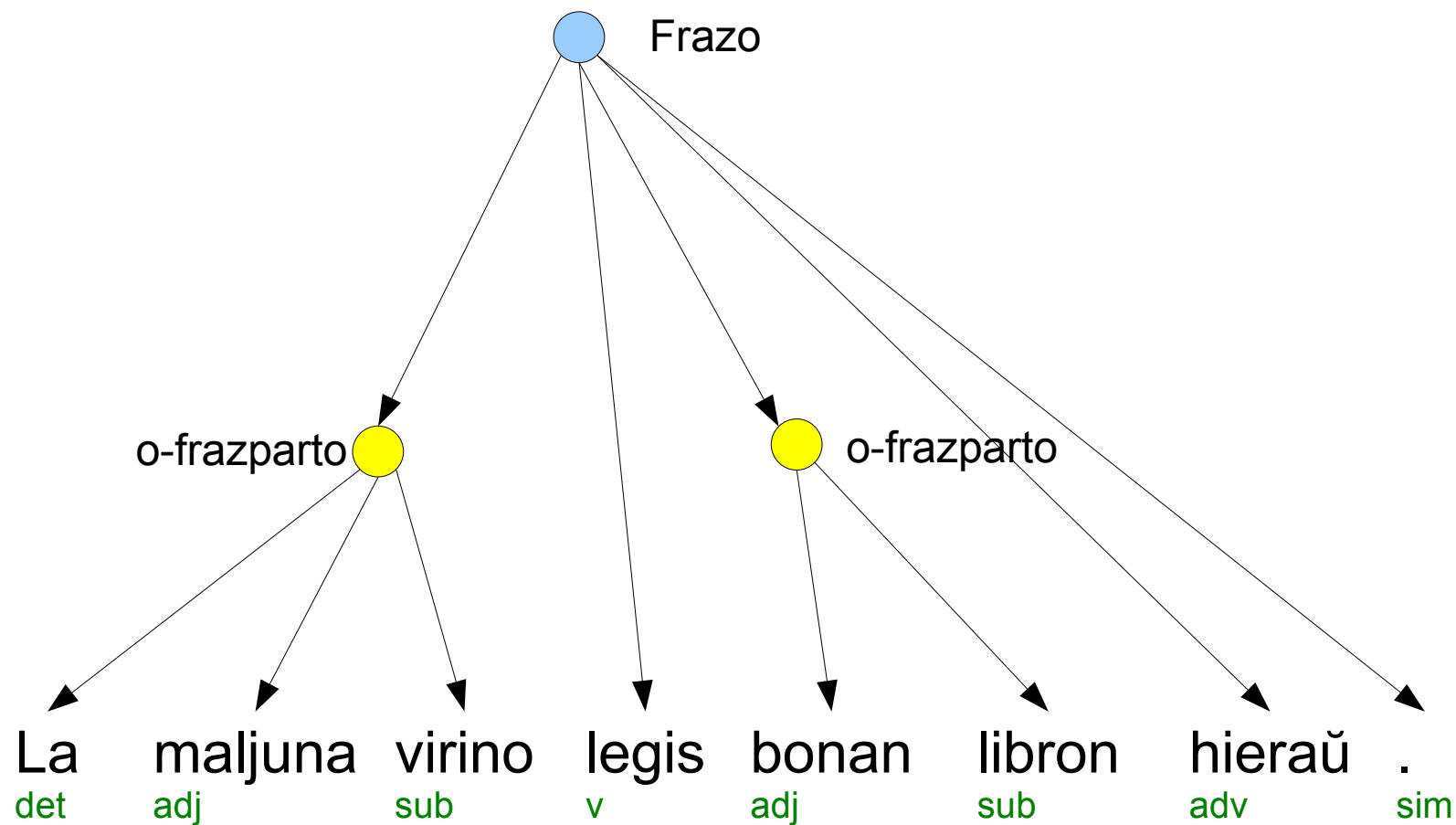
- Vortoj estas folioj
- Branĉoj estas frazpartoj aŭ subfrazoj
- Radikoj estas frazoj

Analizi frazon signifas konstrui la sintaksan arbon de frazo



# Sintaksa Analizo - frazpartoj

Ekzistaj multaj ebloj konstrui sintaksajn arbojn. Esperantilo konstruas ilin de folioj (*Button-Up*)  
Oni angle nomas tiun ŝtupon (*chunking*) – vortgrupigo  
En tiu ŝtupo oni nur pritraktas najbarajn vortojn

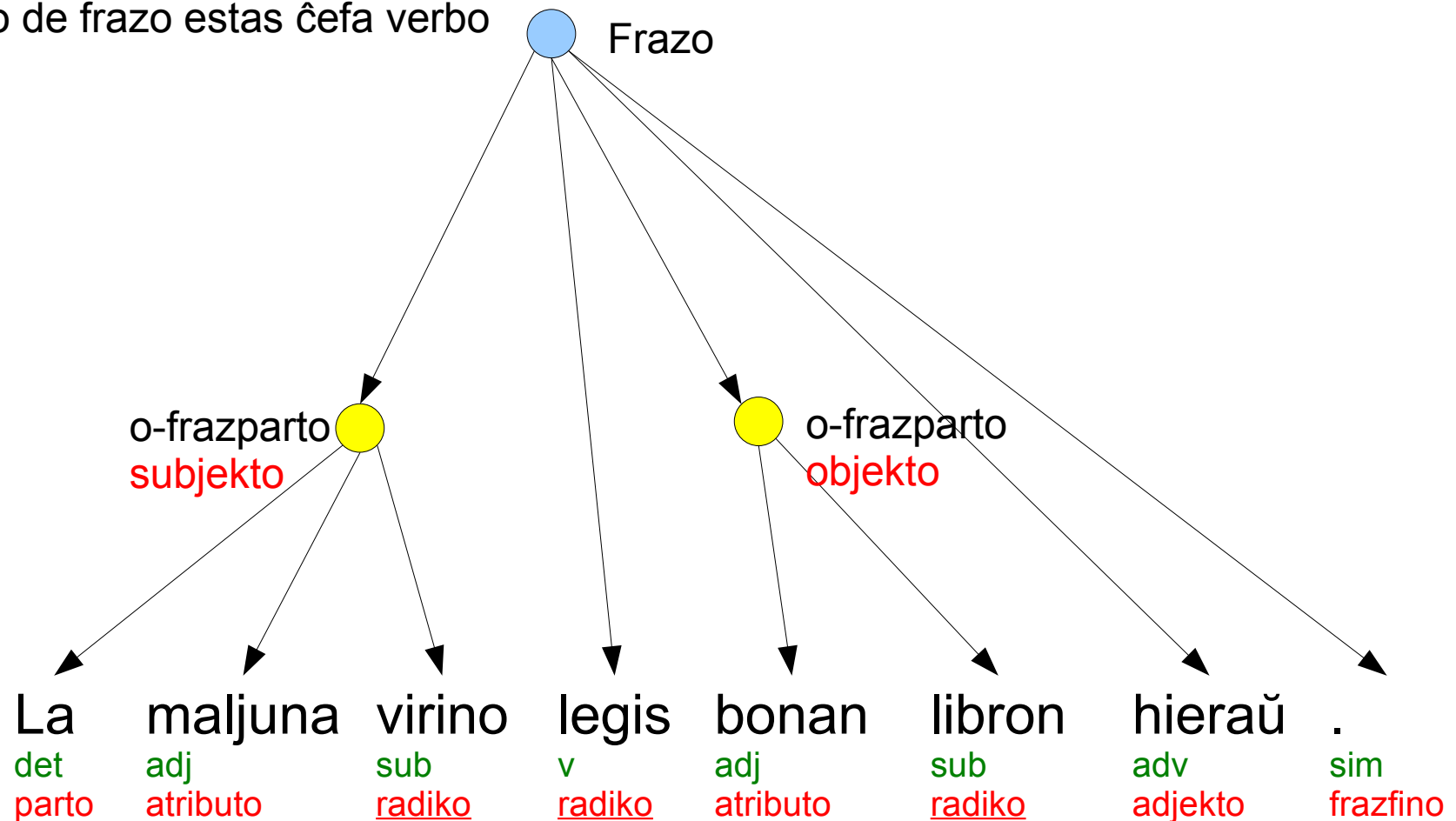




# Sintaksa Analizo – Kategorioj kaj Funkcioj

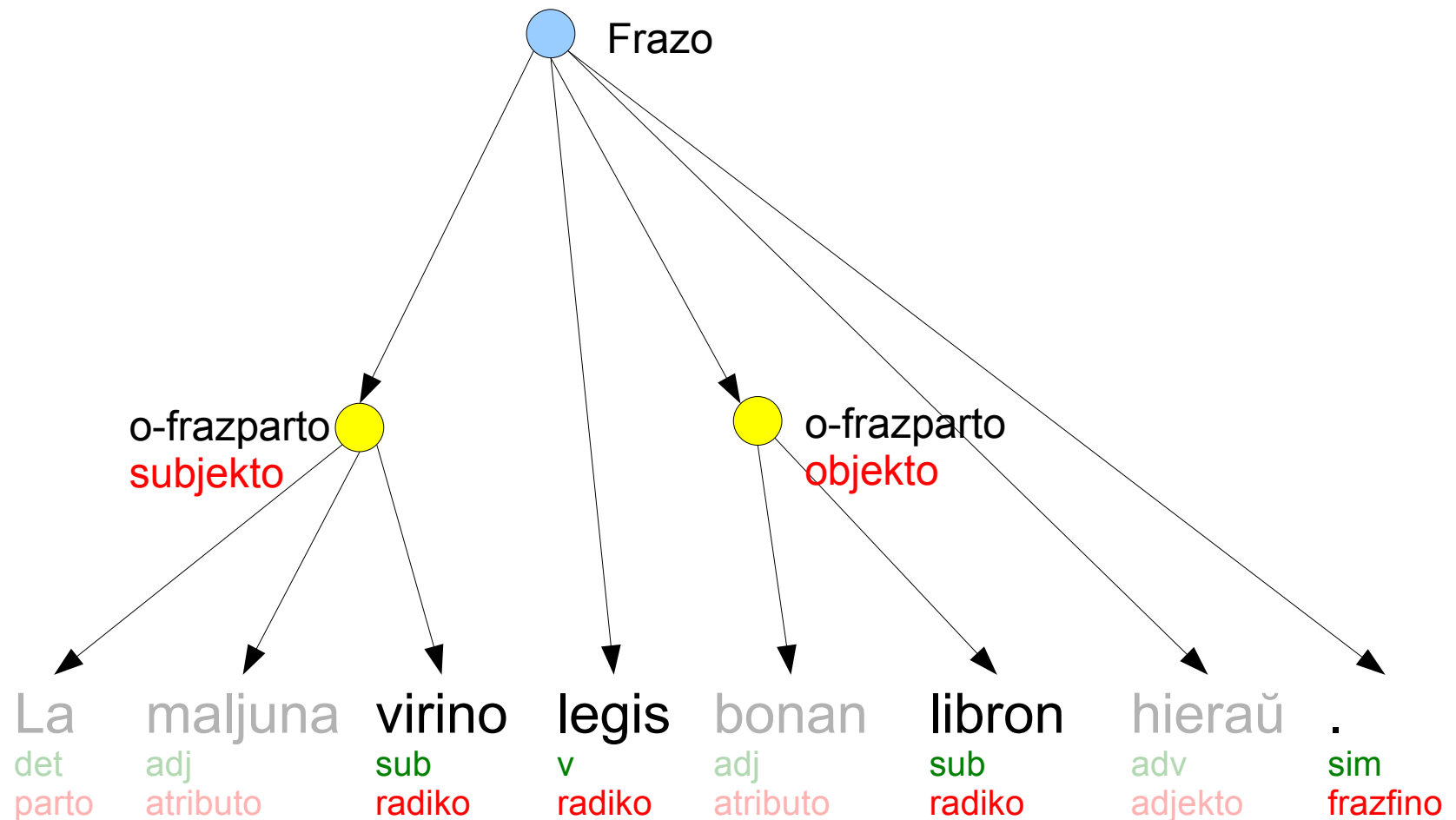
Kutime ĉiu elemento de sintaksa arbo (folio aŭ branĉo) havas funkcion kaj tipon.  
Ĉiu branĉo havas unu radikon.

Radiko de frazo estas ĉefa verbo



# Trajtoj de radikoj

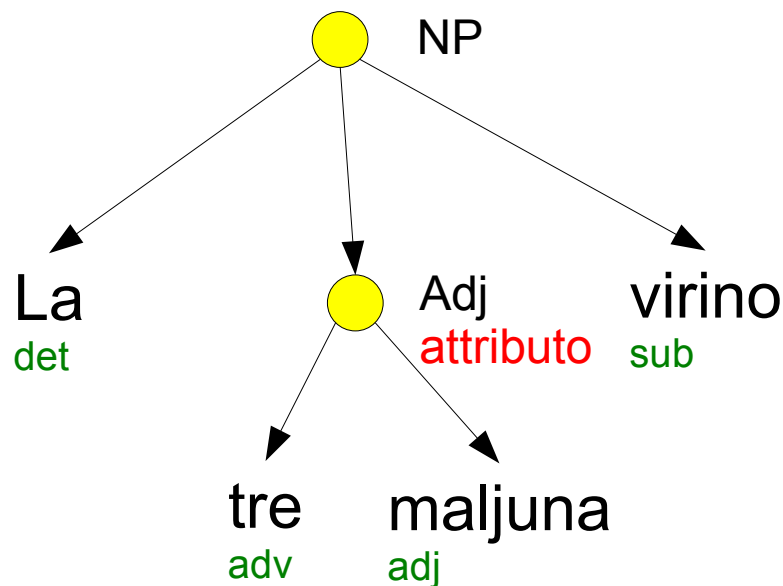
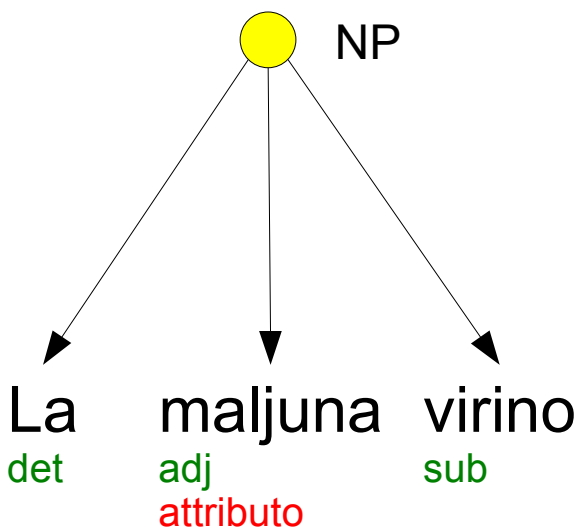
Oni povas forigi ĉiun elementon de branĉo krome radikoj sen perdi la gramatikan korekton



# Branĉoj en Esperantilo - Ekzemploj

Mallongigo	Nomo	Radiko	Ekzemplo	Funkcioj
np	substantiva frazparto	substantivo	bona libro	subjekto, objekto
adj	adjektiva frazparto	adjektivo	tre bona	atributo
adj	adverba frazparto	adverbo	tre ofte	adjekto
nfv	infinitiva verba frazparto	i-verbo	legi librojn	
subsup	substantiva suplemento	substantivo	libro de patro	kiel np
pp	prepozicia	prepozicio	al patro	indirekta objekto, adjekto, ...

Ofte branĉo rolas same kiel la sola radiko



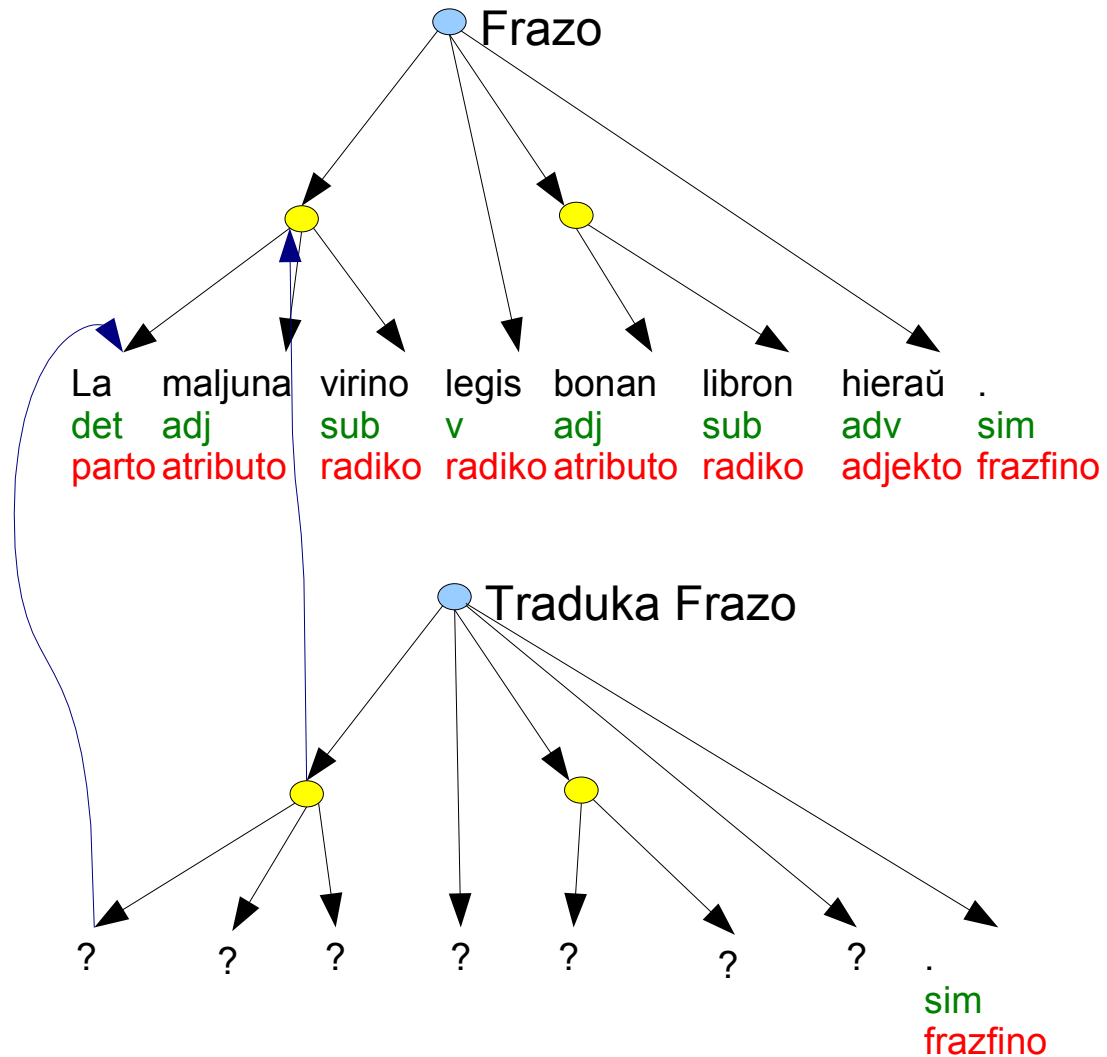
# Tradukado

- Korekta sintaksa analizo estas premiso por bona traduko
- Hipotezoj:
  - sintaksa arbo de traduko kunordas kun sintaksa arbo de fonto
  - tradukado estas transformigo de sintaksaj arboj

# Tradukado - Defioj

- elekto de sinonimoj
- fleksio
- traduko dependa de funkcio
- manko de necesa informo
- ordo
- traduko de idiomaĵoj, idiotismoj

# Teknika proceso de tradukado – kopiado



La sintaksa arbo estas duobligita.  
Tradukaj elemento konas iliajn  
fontajn elementojn.

# Traduko de elementoj – pola lingvo

la	= {}
maljuna	= stary
virino	= kobieta
legi	= czytać (I ndk), przeczytać (I dk), odtwarzać (I ndk - <i>tech</i> )
bona	= dobry
libro	= książka 112*
hieraŭ	= wczoraj

Rimarku, ke el vortaro devenas aldonaj leksikaj informoj pri fleksio kaj gramatiko

112\* - Estas tiel nomata fleksia grupo por polaj substantivoj. ĝi specifas la tabelon kun finaĵoj.

„la“ – ne havas polan tradukon, ĉar pola lingvo ne konas determinilon

„ndk“ – (nie dokonany) signifas aspekton de verbo; ago ne estas finita

# Fleksio en pola lingvo

Pola lingvo, kiel ĉiu slava lingvo, havas tre riĉan kaj malregulan fleksion

## Substantivoj - libro

M	książka	M	książki
D	książki	D	książek
C	książce	C	książkom
B	książkę	B	książki
N	książką	N	książkami
Ms	książce	Ms	książkach
W	książko	W	książki

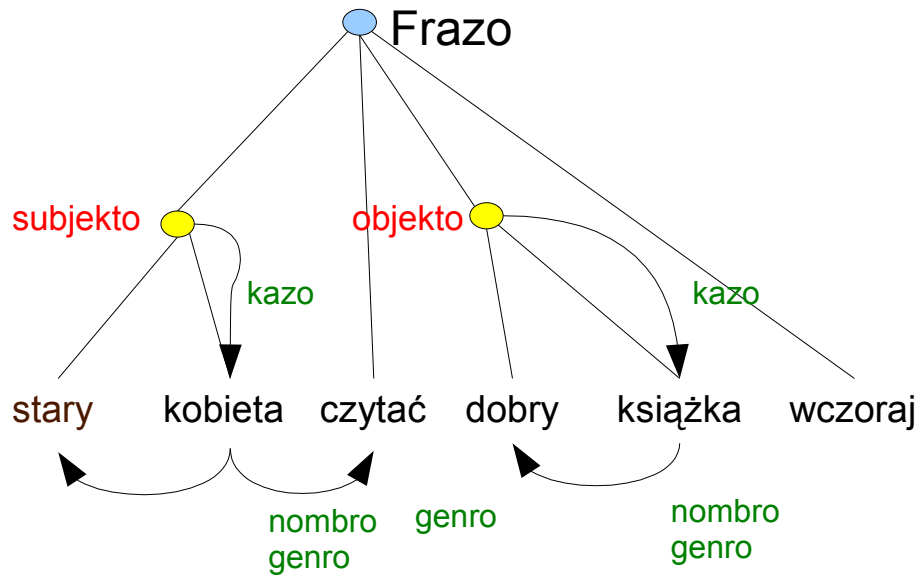
## Verbo - legi

1	czytam	(ms) czytałem (ze) czytałam
2	czytasz	(ms) czytałeś (ze) czytałaś
3	czyta	(on) czytał (ona) czytała (ono) czytało
1	czytamy	(ms) czytaliśmy (ze) czytałyśmy
2	czytacie	(ms) czytaliście (ze) czytałyście
3	czytają	oni czytali one czytały

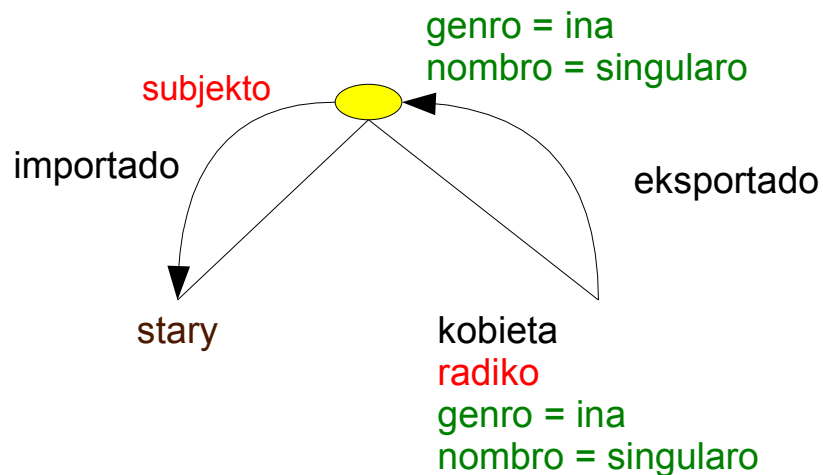
Fleksio	Substantivo	Adjektivo	Verbo	Numeralo	
nombro	X	X	X	X	2 nombroj
kazo	X	X	X	X	7 kazoj
genro	X	X	X	X	6 genroj
tempo			X		3 tempoj



# Fleksio

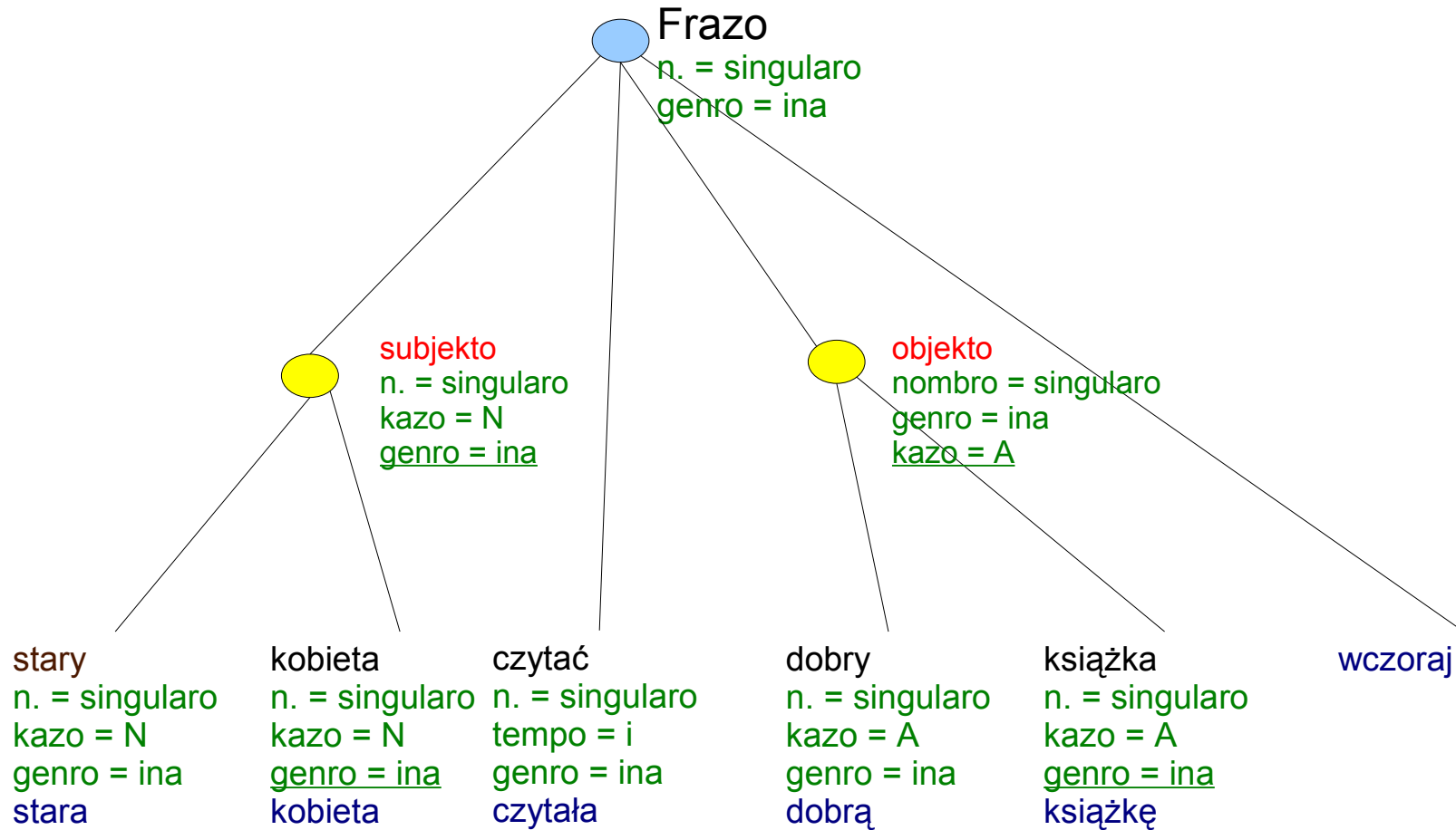


- La necesaj informoj por fleksio devenas de:
- originala elemento (ekz. tempo kaj nombro)
  - radika elemento de parenco
  - funkcio (objekto = akuzativo = Biernik)
  - subjekto (genro por verbo)



- Esperantilo uzas sekvan proceson:
- Radikaj elementaj eksportas la fleksiajn informojn al parencoj
  - Neradikaj elementoj importas la fleksiajn informojn de parencoj

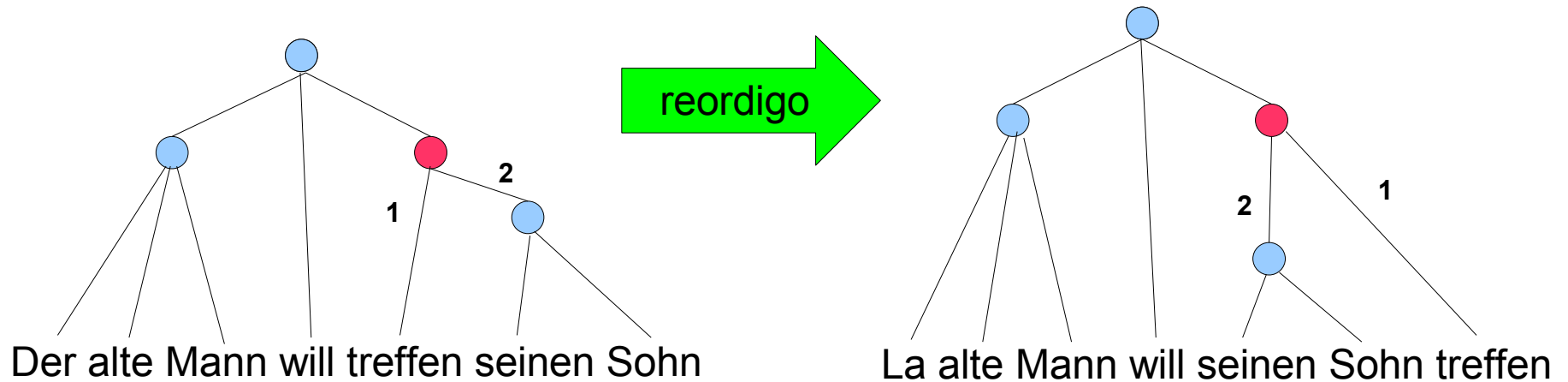
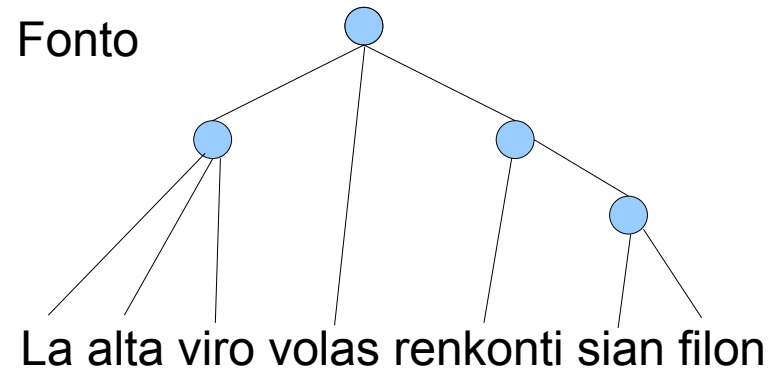
# Plenumo de fleksio



# Reordigo

- multaj lingvo havas ridigan ordon de vortoj en frazo ekz.:  
subjekto + verbo + objekto
- kelkaj lingvoj havas preferataj aŭ kutiman ordon de vortoj
- ĉar frazpartoj ne ŝanĝiĝas ofte oni devas nur ŝanĝi la ordon de infanoj por unu asto


# Ordo – germana lingvo



Nur ruĝa branĉo devas ŝanĝi la ordon de infanoj

# Mankantaj informoj

Esperanta gramatiko pro malriĉa fleksio ofte ne enhavas informojn necesajn por tradukado

Vi legis	 eblaj tradukoj	Gerno	Nombro	Traduko
		ina	singularo	Ty czytałaś
		vira	pluralo	Ty czytałeś
		ina	pluralo	Wy czytałyście
		vira	singularo	Wy czytaliście

eblaj solvoj

- interaktiva tradukado (homo liveras informoj per dialogo kun komputilo)
- preferaĵoj
- aldona profunda analizo

# Tradukado de idiomaĵoj

- Por pli bona traduko. Programo provas traduki tutajn frazpartojn. Por tiu estas necesa alia traduka vortaro

perdi la vojon = **zabłądzić**

esti prava = **mieć rację** (*havi rajton*)

Ankaŭ tiu tradukado povas baziĝi sur sintaksaj arboj

# Fino

- Vere, la tradukado estas pli komplika
- Komputilo laboras surbaze de grandaj faciligoj
- ĉiu lingvo havas propran regulojn

Libera senkosta traduka programo

**[www.esperantilo.org](http://www.esperantilo.org)**