

SL0720Y

séance 1 : étiquetage morpho-syntaxique

Franck Sajous - CLLE-ERSS

`sajous@univ-tlse2.fr`

<http://w3.erss.univ-tlse2.fr/membre/fsajous/>

10 février 2011

Table des matières

1	À la découverte de Treetagger	2
2	Installation	2
3	Utilisation	3
3.1	tag-french	3
3.2	Le programme tree-tagger	5
4	Jeu d'étiquettes	5
4.1	Pour le français	5
4.2	Pour l'anglais	6
4.3	Un autre étiqueteur : Cordial analyseur	6
5	Étiquetage des mots inconnus	8
5.1	Études de cas avec TreeTagger	8
5.2	Affichage des probabilités	8
5.3	Cordial	8
A	Syntaxe d'appel de TreeTagger	10
A.1	Syntaxe TreeTagger	10
A.2	Sortie dans un fichier	10
B	Jeu d'étiquettes pour le Français	11
C	Jeu d'étiquettes pour l'Anglais : Penn Treebank Tagset	12
D	<i>Le dormeur du val</i>, Arthur Rimbaud, 1870	13
E	<i>Le dormeur du val</i> étiqueté par TreeTagger	13
F	<i>Le dormeur du val</i> étiqueté par Cordial analyseur	15
G	Pour installer Treetagger chez vous	17

Note : toutes les urls données dans ce document sont présentes sous forme de liens cliquables à l'adresse :

<http://w3.erss.univ-tlse2.fr/membre/fsajous/SDL/SL0720Y/1/>

1 À la découverte de Treetagger

1. Répertoire de travail

À la racine du disque C: doit se trouver un répertoire nommé SL0720. Si ce n'est pas le cas, créez-le : ouvrez le poste de travail, clic droit sur C:, puis *ouvrir*. Clic droit dans la fenêtre, puis *Nouveau/dossier*. C'est dans ce répertoire que vous effectuerez vos manipulations et que vous installerez TreeTagger (section 2).

- Ouvrez l'éditeur de texte Textpad (icône sur le bureau ou menu *Démarrer/Tous les programmes*) ou Wordpad (menu *Démarrer/Programmes/Accessoires*).
- Saisissez quelques phrases et enregistrez le document dans le répertoire C:\SL0720.
- Ouvrez Firefox et saisissez l'url ci-dessous dans la barre d'adresse :
<http://cental.fltr.ucl.ac.be/treetagger/>
(ou cherchez *treetagger cental* dans un moteur de recherche);
- Cliquez sur le bouton *Parcourir* en face de *texte à étiqueter*, puis sélectionnez le fichier texte que vous avez créé dans la boîte de dialogue qui s'affiche et validez.
- Cliquez sur envoyer (ou *submit query*).
- Cliquez sur le lien *Télécharger le fichier étiqueté* (clic gauche pour l'afficher dans le navigateur).
- Observez le résultat : de quelles informations dispose-t-on en plus du texte initial ?

2 Installation

1. Téléchargement de TreeTagger

Ouvrez un navigateur web et saisissez dans la barre d'adresse l'url :

<http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>
ou cherchez "*treetagger*" dans un moteur de recherche.

Trouvez le lien *Windows version* (entre les sections *Parameter files* et *tagsets*, ou Ctrl+F et *Windows version* dans la barre de recherche) et effectuez un clic droit, puis sélectionnez *enregistrer la cible du lien sous* dans le menu contextuel. Enregistrez le fichier *tree-tagger-windows-3.1.zip* dans le répertoire C:\SL0720.

2. Extraction de l'archive TreeTagger

Effectuez un clic droit sur le fichier *.zip* et sélectionnez dans le menu contextuel *7-zip/extract here*.

Le répertoire C:\SL0720 contient à présent un sous-répertoire *TreeTagger* qui contient lui-même les sous-répertoires *bin*, *cmd* et *lib*.

3. Téléchargement des fichiers de paramètres

Pour utiliser TreeTagger avec une langue donnée, il faut lui fournir un fichier de paramètres correspondant. Dans la section *Parameter files for PC*, effectuez un clic droit sur les liens *English parameter file* et *French parameter file (Latin1)* et enregistrez les fichiers dans `C:\SL0720\TreeTagger\lib`.

4. Extraction des fichiers de paramètres

Le répertoire `C:\SL0720\TreeTagger\lib` doit contenir les fichiers suffixés `.bin.gz`. Décompressez chacun de ces fichiers par un clic droit, puis *7-zip/extract here*. Renommez-les (clic-droit, puis *Renommer*) respectivement en `french.par` et `english.par`.

5. Modification du chemin d'exécution

Pour exécuter un programme, on peut :

- soit saisir son chemin (endroit où il se trouve sur le disque dur) complet (on dit *absolu*) ;
- soit saisir simplement son nom. Dans ce cas, il faut que le système d'exploitation (ici, Windows) sache où trouver le programme. La *variable d'environnement* `PATH` sert à spécifier la liste des répertoires où se trouvent les programmes exécutables que nous souhaitons invoquer uniquement par leurs noms. Nous allons ajouter le répertoire de TreeTagger à cette variable :
 - (a) cliquez sur le bouton **Démarrer** de la barre des tâches, puis **Paramètres** et **Panneau de configuration** (cf fig. 1(a)) ;
 - (b) double-cliquez sur l'icône **Système** (cf fig. 1(b)) ;
 - (c) cliquez sur l'onglet **Avancé** (en haut à droite) puis sur le bouton **Variables d'environnement** (en bas à gauche, cf fig. 1(c)) ;
 - (d) cliquez sur le bouton **Nouveau** de la partie **Variables utilisateur** (cf fig 1(d)) ;
 - (e) ajoutez la variable `PATH` avec la valeur `C:\SL0720\TreeTagger\bin` (cf fig 1(e)) ;
 - (f) validez en appuyant sur **OK** dans les fenêtres restées ouvertes.

6. Modification des scripts `tag-french` et `tag-english`

Ouvrez le fichier `C:\SL0720\TreeTagger\bin\tag-french.bat` (clic droit et *modifier*) et changez la ligne :

```
set TAGDIR=C:\TreeTagger
```

en :

```
set TAGDIR=C:\SL0720\TreeTagger
```

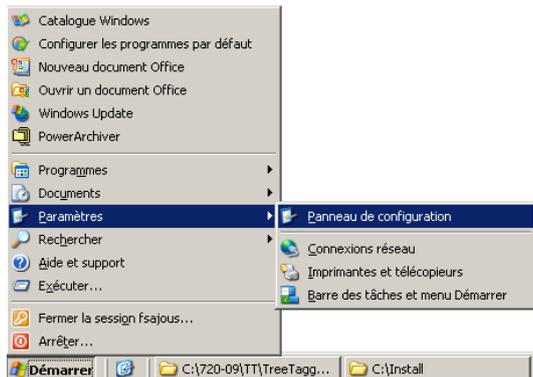
Enregistrez le fichier (menu *Fichier/Enregistrer* ou `Ctrl+S`) et fermez-le (menu *Fichier/Quitter* ou `Alt+F4`).

Faites de même avec le fichier `tag-english.bat`.

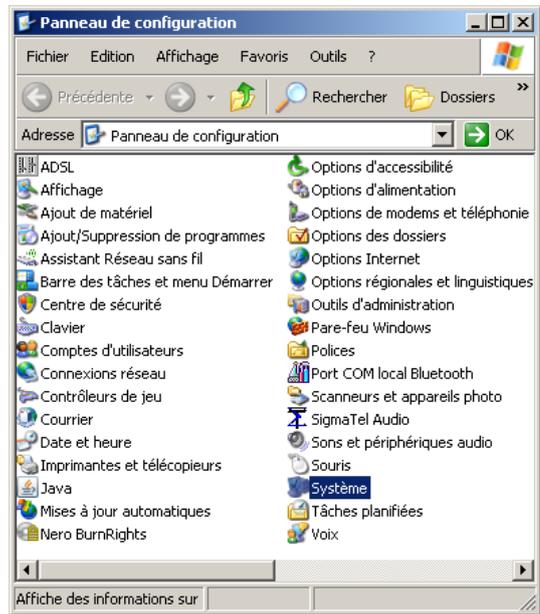
3 Utilisation

3.1 tag-french

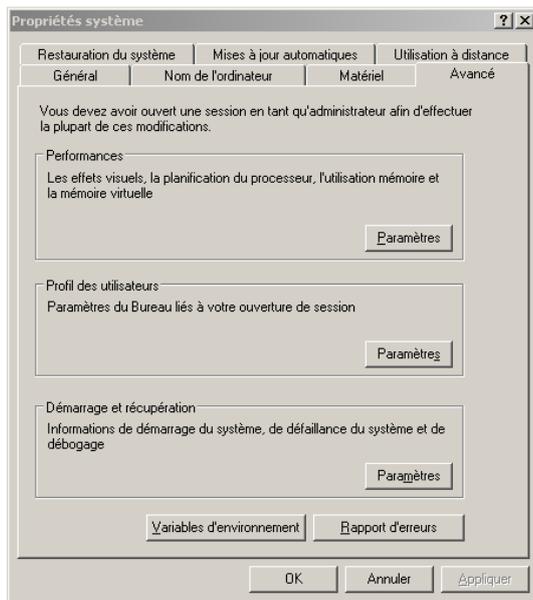
1. créez un fichier texte dans le dossier `C:\SL0720` :



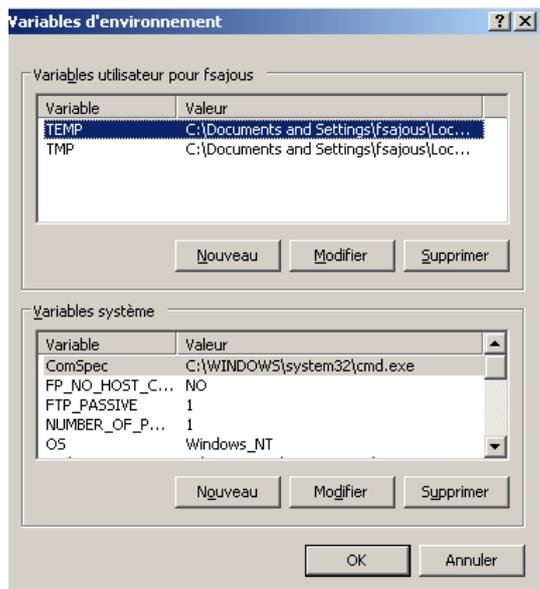
(a) Ouverture du panneau de configuration



(b) Icône Système



(c) Bouton variables d'environnement



(d) Variables utilisateurs



(e) Nouvelle variable PATH

FIG. 1 – Modification de la variable d'environnement PATH

- soit ouvrez ce dossier dans l’explorateur, puis menu **Fichier/Nouveau/Document texte** et saisissez un nom de fichier (eg. `testTT.txt`), validez et double-cliquez dessus ;
 - soit ouvrez Textpad ou Wordpad et enregistrez le fichier dans le répertoire.
2. saisissez quelques phrases (au moins deux) dans ce fichier et enregistrez-le.
 3. ouvrez une console DOS : cliquez sur le bouton **Démarrer** de la barre des tâches, puis :
 - soit cliquez sur **Exécuter** et saisissez `cmd`, puis validez ;
 - soit allez dans le menu **Programme/Accessoires** et cliquez sur **Invite de commande**.
 4. allez dans le répertoire où se trouve le fichier texte :


```
cd C:\SL0720\
```
 5. puis lancez TreeTagger sur ce fichier :


```
tag-french monfichier.txt
```

Note : le fait d’avoir positionné la variable `PATH` nous dispense de saisir le chemin d’accès au programme. Sinon, il aurait fallu saisir :

```
C:\SL0720\TreeTagger\bin\tag-french monfichier.txt
```

Quelles sont toutes les fonctions effectuées par ce programme ?

3.2 Le programme tree-tagger

Le script `tag-french` vu précédemment permet de lancer rapidement l’étiquetage d’un texte en français avec des options spécifiques, non modifiables. Il fait appel au programme `tree-tagger`, qui est utilisé pour toutes les langues traitées. Seul diffère le fichier de paramètres (propre à chaque langue) qu’il prend en argument.

La syntaxe de ce programme est la suivante (pour la syntaxe détaillée, voir l’annexe A) :

```
tree-tagger [options] fichier_de_paramètres fichier_texte
```

Lancez le programme sur le fichier précédent :

```
tree-tagger C:\SL0720\TreeTagger\lib\french.par monfichier.txt
```

Que constatez-vous ?

Éditez le fichier texte et séparez chaque item (mot ou signe de ponctuation) par un retour-chariot (retour à la ligne), enregistrez et relancez la commande ci-dessus. Qu’observez-vous ?

Attention : lors de la séparation manuelle par des retour-chariots, veillez à ne laisser de blancs (espaces) ni avant, ni après les mots.

À l’aide de la syntaxe donnée dans l’annexe A, obtenez une sortie telle que celle obtenue section 1 (*i.e.* avec la forme du mot et son lemme).

4 Jeu d’étiquettes

4.1 Pour le français

1. récupérez sur la page du TD le fichier `gloubiboulga.txt`¹

¹Texte pris dans l’article Wikipédia <http://fr.wikipedia.org/wiki/Gloubi-boulga>

2. étiquetez ce fichier (forme, partie du discours, lemme).
 - (a) qu'observe-t-on dans la colonne des lemmes?
 - (b) quelles sont les étiquettes pour les marques de ponctuation?
 - (c) à quoi correspondent-elles?

4.2 Pour l'anglais

Récupérez le fichier `holy_grail.txt`² et lancez le script `tag-english` :
`tag-english holy_grail.txt`

Quelle remarque concernant les marques de ponctuation peut-on faire par rapport au français ?

Quelles autres remarques peut-on faire concernant le jeu d'étiquettes ?

4.3 Un autre étiqueteur : Cordial analyseur

Les exemples ci-après donnent une comparaison entre les étiquetages de TreeTagger et de Cordial Analyseur (Synapse Développement) sur *Le dormeur du val* d'A. Rimbaud³.

1. Les deux premiers vers étiquetés par **Cordial** sont représentés figure 2. Quelles informations **Cordial** apporte-t-il en plus de **TreeTagger** ?
2. Les versions étiquetées (forme, lemme et partie du discours) par TreeTagger et Cordial analyseur du *dormeur du val* sont données respectivement dans les annexes E et F. Quelle(s) différence(s) observez-vous ?
3. Relevez des erreurs d'étiquetage (ou résultats inattendus) pour chaque étiqueteur et comparez avec l'étiquetage de l'autre.

²Extrait issu du site *Unofficial Monty Python Home Page* :
<http://www.muscomp.com/python.html>

³Cette comparaison est empruntée à Habert 2005, dans *Instruments et ressources électroniques pour le français*; voir la ref. bibliographique sur la page du module.

```

===== DEBUT DE PHRASE =====
1 C' ce PDS Pd-.n 1|1 S 1 est Principale
2 est être VINDP3S Vmip3s 2 V 1 est Principale
3 un un DETMS Da-ms-i 4|4 B 1 est Principale
4 trou trou NCMS Ncms 4|4 B 1 est Principale
5 de de PREP Sp 6|4 B 1 est Principale
6 verdure verdure NCFS Ncfs 6|4 B 1 est Principale
7 où où PRI Ptr-.- 7|7 - 2 chante Relative
8 chante chanter VINDP3S Vmip3s 8 V 2 chante Relative
9 une un DETIFS Da-fs-i 10|10 T 2 chante Relative
10 rivière rivière NCFS Ncfs 10|10 T 2 chante Relative
===== FIN DE PHRASE =====

===== DEBUT DE PHRASE =====
1 Accrochant accrocher VPARPRES Vmpp-- 1 - 1 Indépendante suspendre/hang,suspend
2 follement follement ADV Rgp - 1 Indépendante
3 aux à le DETDP3G Da-p-d 4|4 - 1 Indépendante
4 herbes herbe NCFP Ncfp 4|4 - 1 Indépendante plante/grass
5 des de le DETDP3G Da-p-i 6|4 - 1 Indépendante
6 haillons haillon NCMP Ncmp 6|4 - 1 Indépendante
===== FIN DE PHRASE =====
ouverture/opening, unobstructed space
pour les oiseaux/sing, warble

```

FIG. 2 – *Le dormeur du val* étiqueté par Cordial (deux premiers vers)

5 Étiquetage des mots inconnus

5.1 Études de cas avec TreeTagger

Dans la partie 4.1, comment expliquer le bon étiquetage des mots inconnus ou mal orthographiés ?

Étiquetez la phrase suivante :

Ce été a été le plus bel été qu'il ait été depuis l'été 2000.

Étiquetez séparément les séquences suivantes :

1. *gnagnagneur*
2. *gnagnanir*
3. *un grand gnagnanir*
4. *gnagnavient*
5. *un grand gnagnavient*

Que peut-on en déduire de cette expérience quant à l'étiquetage des mots inconnus ?

Étiquetez séparément les séquences suivantes :

1. *guerre civile*
2. *guerre cccivile*
3. *guerre civilation*

Quelle nouvelle supposition peut-on faire ?

5.2 Affichage des probabilités

Trouvez dans l'annexe A la syntaxe d'appel de TreeTagger pour afficher les différentes étiquettes possibles pour un même mot et étiquetez à nouveau les séquences de la section 5.1.

5.3 Cordial

Ci-dessous sont donnés des résultats d'analyse par Cordial pour les séquences de la section 5.1. Commentez-les.

==== DEBUT DE PHRASE ====	==== DEBUT DE PHRASE ====
Cet ce DETDEM	Un un DETIMS
été été NCMS	grand grand ADJMS
a avoir VINDP3S	gnagnanir gnagnanir NCI
été être VPARPMS	===== FIN DE PHRASE =====
le plus le plus ADJIND	==== DEBUT DE PHRASE ====
bel bel ADJMS	gnagnagir gnagnagir NCI
été être VPARPMS	qui qui PRI
qu' que SUB	roule rouler VINDP3S
il il PPER3S	n' ne ADV
ait avoir VSUBP3S	amasse amasser VIMPP2S
été être VPARPMS	pas pas ADV
depuis depuis PREP	mousse mousse ADJSIG
l' le DETDMS	. . PCTFORTE
été été NCMS	===== FIN DE PHRASE =====
2000 2000 NCMIN	==== DEBUT DE PHRASE ====
. . PCTFORTE	Il il PPER3S
===== FIN DE PHRASE =====	est être VINDP3S
==== DEBUT DE PHRASE ====	interdit interdit ADJMS
guerre guerre NCFS	de de PREP
ccivile ccivile ADJINV	gnagnanir gnagnanir NCI
===== FIN DE PHRASE =====	dans dans PREP
==== DEBUT DE PHRASE ====	les le DETDPIG
guerre guerre NCFS	lieux lieu NCMP
civilation civilation ADJINV	publics public ADJMP
===== FIN DE PHRASE =====	. . PCTFORTE
==== DEBUT DE PHRASE ====	===== FIN DE PHRASE =====
gnagnaneur gnagnaneur NCI	
===== FIN DE PHRASE =====	
==== DEBUT DE PHRASE ====	
gnagnanir gnagnanir NCI	
===== FIN DE PHRASE =====	

FIG. 3 – Étiquetage des mots inconnus par Cordial analyseur

A Syntaxe d'appel de TreeTagger

A.1 Syntaxe TreeTagger

USAGE: tree-tagger {-options-} <parameter file> {<input file> {<output file>}}

OPTIONS:

- token: Print the token
- lemma: Print the lemma
- sgml: Don't tag SGML annotations
- threshold <p>: Print all tags of a word with a probability higher than <p> times the largest probability
- prob: Print tag probabilities
- ignore-prefix: Ignore prefix when guessing pos for unknown words.
- no-unknown: Print the token rather than <unknown> for unknown lemmas
- no-heuristics: Don't use any heuristics in the lexicon
- no-hyphen-heuristic: Turn off the heuristic for guessing the parts of speech of unknown hyphenated words
- quiet: Don't print status messages
- pt-with-lemma: pretagging with lemmata
- pt-with-prob: pretagging with probabilities
- files <f>: read names of input and output files pairwise from <f>
- eos-tag <tag>: The SGML tag <tag> signals the end of a sentence.
This option implies the option -sgml

Some more exotic options:

- proto: Print lexical information for each word
- gramotron: Same as -proto but with a different format
- proto-with-prob: Same as -proto but with lexical tag probabilities
- eps <epsilon>: Set minimal tag frequency to <epsilon>
- base: Use only lexical probabilities for tagging
- print-prob-tree: Print the transition probability tree and exit
- print-suffix-tree: Print the suffix probabilities and exit

A.2 Sortie dans un fichier

Par défaut, TreeTagger écrit ses résultats sur la *sortie standard* : dans notre cas, il s'agit de la console DOS depuis laquelle il est invoqué. Pour écrire les résultats de TreeTagger dans un fichier (on dit *rediriger la sortie de TreeTagger*), la syntaxe est la suivante :

```
tree-tagger {-options-} <parameter_file> <input_file> > fichier_sortie
```

Il suffit de faire suivre la ligne de commande par un chevron fermant et le nom de fichier dans lequel on veut stocker les résultats.

B Jeu d'étiquettes pour le Français

ABR	abreviation
ADJ	adjective
ADV	adverb
DET:ART	article
DET:POS	possessive pronoun (ma, ta, ...)
INT	interjection
KON	conjunction
NAM	proper name
NOM	noun
NUM	numeral
PRO	pronoun
PRO:DEM	demonstrative pronoun
PRO:IND	indefinite pronoun
PRO:PER	personal pronoun
PRO:POS	possessive pronoun (mien, tien, ...)
PRO:REL	relative pronoun
PRP	preposition
PRP:det	preposition plus article (au, du, aux, des)
PUN	punctuation
PUN:cit	punctuation citation
SENT	sentence tag
SYM	symbol
VER:cond	verb conditional
VER:futu	verb futur
VER:impe	verb imperative
VER:impf	verb imperfect
VER:infi	verb infinitive
VER:pper	verb past participle
VER:ppre	verb present participle
VER:pres	verb present
VER:simp	verb simple past
VER:subi	verb subjunctive imperfect
VER:subp	verb subjunctive present

C Jeu d'étiquettes pour l'Anglais : Penn Treebank Tagset

CC	Coordinating conjunction
CD	Cardinal number
DT	Determiner
EX	Existential there
FW	Foreign word
IN	Preposition or subordinating conjunction
JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, superlative
LS	List item marker
MD	Modal
NN	Noun, singular or mass
NNS	Noun, plural
NP	Proper noun, singular
NPS	Proper noun, plural
PDT	Predeterminer
POS	Possessive ending
PP	Personal pronoun
PP\$	Possessive pronoun
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative
RP	Particle
SYM	Symbol
TO	to
UH	Interjection
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, gerund or present participle
VBN	Verb, past participle
VBP	Verb, non-3rd person singular present
VBZ	Verb, 3rd person singular present
WDT	Wh-determiner
WP	Wh-pronoun
WP\$	Possessive wh-pronoun
WRB	Wh-adverb

D *Le dormeur du val*, Arthur Rimbaud, 1870

- 1 C'est un trou de verdure où chante une rivière
 2 Accrochant follement aux herbes des haillons
 3 D'argent ; où le soleil , de la montagne fière ,
 4 Luit : c'est un petit val qui mousse de rayons.
- 5 Un soldat jeune , bouche ouverte , tête nue ,
 6 Et la nuque baignant dans le frais cresson bleu ,
 7 Dort ; il est étendu dans l'herbe , sous la nue ,
 8 Pâle dans son lit vert où la lumière pleut .
- 9 Les pieds dans les glaïeuls , il dort . Souriant comme
 10 Sourirait un enfant malade , il fait un somme :
 11 Nature , berce-le chaudement : il a froid .
- 12 Les parfums ne font pas frissonner sa narine ;
 13 Il dort dans le soleil , la main sur sa poitrine
 14 Tranquille . Il a deux trous rouges au côté droit .

E *Le dormeur du val* étiqueté par TreeTagger

C'	PRO:DEM	ce	,	PUN	,
est	VER:pres	être	Luit	VER:pres	luire
un	DET:ART	un	:	PUN	:
trou	NOM	trou	c'	PRO:DEM	ce
de	PRP	de	est	VER:pres	être
verdure	NOM	verdure	un	DET:ART	un
où	PRO:REL	où	petit	ADJ	petit
chante	VER:pres	chanter	val	NOM	val
une	DET:ART	un	qui	PRO:REL	qui
rivière	NOM	rivière	mousse	NOM	mousse
Accrochant	VER:ppre	accrocher	de	PRP	de
follement	ADV	follement	rayons	NOM	rayon
aux	PRP:det	au	.	SENT	.
herbes	NOM	herbe	Un	DET:ART	un
des	PRP:det	du	soldat	NOM	soldat
haillons	NOM	haillon	jeune	ADJ	jeune
D'	PRP	de	,	PUN	,
argent	NOM	argent	bouche	NOM	bouche
;	PUN	;	ouverte	ADJ	ouvert
où	PRO:REL	où	,	PUN	,
le	DET:ART	le	tête	NOM	tête
soleil	NOM	soleil	nue	NOM	nue
,	PUN	,	,	PUN	,
de	PRP	de	Et	KON	et
la	DET:ART	le	la	DET:ART	le
montagne	NOM	montagne	nuque	NOM	nuque
fière	ADJ	fier	baignant	VER:ppre	baigner

dans	PRP	dans		fait	VER:pres	faire
le	DET:ART	le		un	DET:ART	un
frais	ADJ	frais		somme	NOM	somme
cresson	NOM	cresson		:	PUN	:
bleu	ADJ	bleu		Nature	NOM	nature
,	PUN	,		,	PUN	,
Dort	VER:pres	dormir		berce	VER:subp	bercer
;	PUN	;		-le	PRO:PER	le
il	PRO:PER	il		chaudement	ADV	chaudement
est	VER:pres	être		:	PUN	:
étendu	VER:pper	étendre		il	PRO:PER	il
dans	PRP	dans		a	VER:pres	avoir
l'	DET:ART	le		froid	ADJ	froid
herbe	NOM	herbe		.	SENT	.
,	PUN	,		Les	DET:ART	le
sous	PRP	sous		parfums	NOM	parfum
la	DET:ART	le		ne	ADV	ne
nue	NOM	nue		font	VER:pres	faire
,	PUN	,		pas	ADV	pas
Pâle	ADJ	pâle		frissonner	VER:infi	frissonner
dans	PRP	dans		sa	DET:POS	son
son	DET:POS	son		narine	NOM	narine
lit	NOM	lit		;	PUN	;
vert	ADJ	vert		Il	PRO:PER	il
où	PRO:REL	où		dort	VER:pres	dormir
la	DET:ART	le		dans	PRP	dans
lumière	NOM	lumière		le	DET:ART	le
pleut	VER:pres	pleuvoir		soleil	NOM	soleil
.	SENT	.		,	PUN	,
Les	DET:ART	le		la	DET:ART	le
pieds	NOM	piéd		main	NOM	main
dans	PRP	dans		sur	PRP	sur
les	DET:ART	le		sa	DET:POS	son
glaieuls	NOM	glaieul		poitrine	NOM	poitrine
,	PUN	,		Tranquille	ADJ	tranquille
il	PRO:PER	il		.	SENT	.
dort	VER:pres	dormir		Il	PRO:PER	il
.	SENT	.		a	VER:pres	avoir
Souriant	VER:ppre	sourire		deux	NUM	deux
comme	ADV	comme		trous	NOM	trou
Sourirait	VER:cond	sourire		rouges	ADJ	rouge
un	DET:ART	un		au	PRP:det	au
enfant	NOM	enfant		côté	NOM	côté
malade	ADJ	malade		droit	ADJ	droit
,	PUN	,		.	SENT	.
il	PRO:PER	il				

F *Le dormeur du val* étiqueté par Cordial analyseur

==== DEBUT DE PHRASE ====	de	de	PREP	
C' ce PDS	rayons	rayon	NCMP	
est être VINDP3S	.	.	PCTFORTE	
un un DETIMS	==== FIN DE PHRASE =====			
trou trou NCMS	==== DEBUT DE PHRASE =====			
de de PREP	Un un DETIMS			
verdure verdure NCFS	soldat soldat NCMS			
où où PRI	jeune jeune ADJSIG			
chante chanter VINDP3S	,	,	PCTFAIB	
une un DETIFS	bouche bouche NCFS			
rivière rivière NCFS	ouverte ouvert ADJFS			
==== FIN DE PHRASE =====	,	,	PCTFAIB	
==== DEBUT DE PHRASE =====	tête nue tête nue			ADV
Accrochant accrocher VPARPRES	,	,	PCTFAIB	
follement follement ADV	==== FIN DE PHRASE =====			
aux à le DETDPIG	==== DEBUT DE PHRASE =====			
herbes herbe NCFP	Et et COO			
des de le DETDPIG	la le DETDFS			
haillons haillon NCMP	nuque nuque NCFS			
==== FIN DE PHRASE =====	baignant baigner VPARPRES			
==== DEBUT DE PHRASE =====	dans dans PREP			
D' de PREP	le le DETDMS			
argent argent ADJINV	frais frais ADJMIN			
;	cresson cresson NCMS			
==== FIN DE PHRASE =====	bleu bleu ADJINV			
==== DEBUT DE PHRASE =====	,	,	PCTFAIB	
où où PRI	==== FIN DE PHRASE =====			
le le DETDMS	==== DEBUT DE PHRASE =====			
soleil soleil NCMS	Dort dormir VINDP3S			
,	;	;	PCTFORTE	
de de PREP	==== FIN DE PHRASE =====			
la le DETDFS	==== DEBUT DE PHRASE =====			
montagne montagne NCFS	il il PPER3S			
fière fier ADJFS	est être VINDP3S			
,	étendu étendre VPARPMS			
==== FIN DE PHRASE =====	dans dans PREP			
==== DEBUT DE PHRASE =====	l' le DETDFS			
Luit luire VINDP3S	herbe herbe NCFS			
:	,	,	PCTFAIB	
==== FIN DE PHRASE =====	sous sous PREP			
==== DEBUT DE PHRASE =====	la le DETDFS			
c' ce PDS	nue nue NCFS			
est être VINDP3S	,	,	PCTFAIB	
un un DETIMS	==== FIN DE PHRASE =====			
petit petit ADJMS	==== DEBUT DE PHRASE =====			
val val NCMS	Pâle pâle ADJSIG			
qui qui PRI	dans dans PREP			
mousse mousser VINDP3S	son son DETPOSS			

lit	lit	NCMS	.	.	PCTFORTE	
vert	vert	ADJMS	=====	FIN DE PHRASE	=====	
où	où	PRI	=====	DEBUT DE PHRASE	=====	
la	le	DETDIFS	Les	le	DETDPIG	
lumière	lumière	NCFS	parfums	parfum	NCMP	
pleut	pleuvoir	VINDP3S	ne	ne	ADV	
.	.	PCTFORTE	font	faire	VINDP3P	
=====	FIN DE PHRASE	=====	pas	pas	ADV	
=====	DEBUT DE PHRASE	=====	frissonner	frissonner	VINF	
Les	le	DETDPIG	sa	son	DETPOSS	
pieds	pied	NCMP	narine	nar	NCFS	
dans	dans	PREP	;	;	PCTFORTE	
les	le	DETDPIG	=====	FIN DE PHRASE	=====	
glaïeuls	glaïeul	NCMP	=====	DEBUT DE PHRASE	=====	
,	,	PCTFAIB	Il	il	PPER3S	
il	il	PPER3S	dort	dormir	VINDP3S	
dort	dormir	VINDP3S	dans	dans	PREP	
.	.	PCTFORTE	le	le	DETDMS	
=====	FIN DE PHRASE	=====	soleil	soleil	NCMS	
=====	DEBUT DE PHRASE	=====	,	,	PCTFAIB	
Souriant	sourire	VPARPRES	la	le	DETDIFS	
comme	comme	SUB	main	main	NCFS	
=====	FIN DE PHRASE	=====	sur	sur	PREP	
=====	DEBUT DE PHRASE	=====	sa	son	DETPOSS	
Sourirait	sourire	VCOMP3S	poitrine	poitrine	NCFS	
un	un	DETIMS	=====	FIN DE PHRASE	=====	
enfant	enfant	NCSIG	=====	DEBUT DE PHRASE	=====	
malade	malade	ADJSIG	Tranquille	Tranquille	NPMS	
,	,	PCTFAIB	.	.	PCTFORTE	
il	il	PPER3S	=====	FIN DE PHRASE	=====	
fait un somme	sourire	VINDP3S	=====	DEBUT DE PHRASE	=====	
:	:	PCTFORTE	Il	il	PPER3S	
=====	FIN DE PHRASE	=====	a	avoir	VINDP3S	
=====	DEBUT DE PHRASE	=====	deux	deux	ADJNUM	
Nature	Nature	NPMS	trous	trou	NCMP	
,	,	PCTFAIB	rouges	rouge	ADJPIG	
berce	bercer	VIMPP2S	au	à le	DETDMS	
-le	le	PPER3S	côté	côté	NCMS	
chaudemment	chaudemment	ADV	droit	droit	ADJMS	
:	:	PCTFORTE	.	.	PCTFORTE	
=====	FIN DE PHRASE	=====	=====	FIN DE PHRASE	=====	
=====	DEBUT DE PHRASE	=====	=====	DEBUT DE PHRASE	=====	
il	il	PPER3S	=====	FIN DE PHRASE	=====	
a froid avoir	VINDP3S					

G Pour installer Treetagger chez vous

En salle machines, la procédure d'installation décrite dans l'énoncé suppose que Perl et 7-zip soient installés. Si ce n'est pas le cas sur votre machine, il faut les télécharger et les installer :

– Perl :

1. Allez sur la page : <http://www.activestate.com/activeperl/downloads/>
2. Cliquez sur l'icône *DOWNLOAD* ou choisir une autre version que celle proposée
3. Enregistrez le fichier et double-cliquez dessus
4. Cliquez sur suivant ou OK autant de fois qu'il le faut

– 7-zip :

1. Allez à l'adresse <http://www.7-zip.org/download.html>
2. Cliquer sur *Download*, en face de *.exe*
3. Enregistrez le fichier et double-cliquez dessus
4. Cliquez sur suivant ou OK autant de fois qu'il le faut