

ISTRUZIONI PROGETTO FASE 1 E 2

Data consegna: 11 Giugno

Questo documento contiene le istruzioni per l'esecuzione della fase 1 e 2 del progetto. Per informazioni o domande scrivere a pennacchiotti@info.uniroma2.it.

Le prime due fasi devono essere eseguite da tutti i membri del gruppo, che dovranno dividere equamente i paragrafi da analizzare tra di loro (ogni paragrafo è quindi analizzato da un singolo annotatore).

1. Preparazione del testo di input

- Accedere la pagina di Wikipedia corrispondente all'indirizzo fornito al gruppo, come indicato in: <http://ai-nlp.info.uniroma2.it/pennacchiotti/teaching/gruppi.html>
- Salvare il file HTML ([nome_frammento].html)
- Copiare la parte testuale della pagina (non il sorgente HTML!), evitando elementi non prettamente testuali (menù, elementi grafici, ecc.), come indicato in *Fig.1*
- Selezionare ogni paragrafo significativo del frammento (come indicato in *Fig.1*) e salvarlo in un file di testo [nome_frammento]_[num_paragrafo].txt.

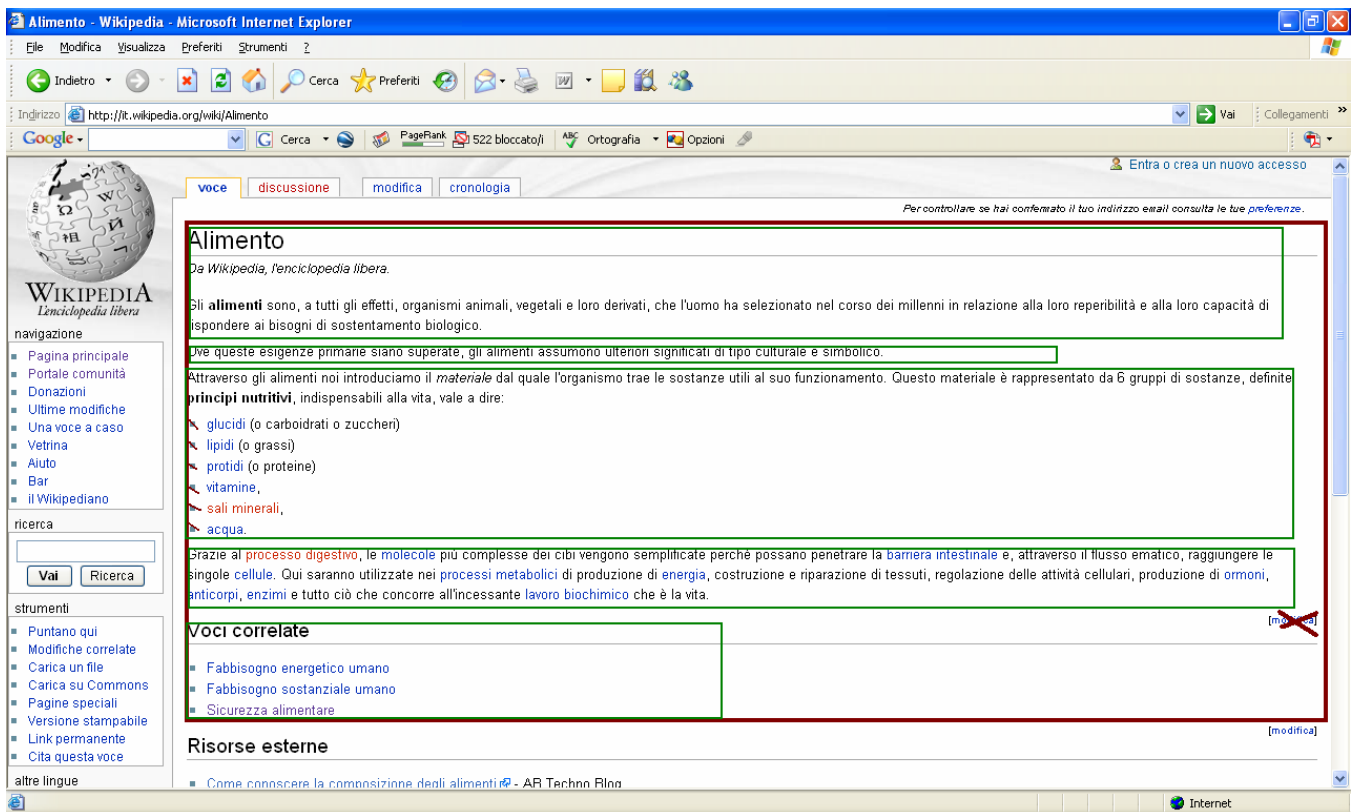


Fig.1. In rosso la sezione del testo da copiare. In verde i diversi paragrafi da salvare.

2. Parsing con Chaos

- Eseguire su ogni file il parsing completo del testo in esso contenuto, utilizzando l'interfaccia grafica di Chaos (`chaosgui.bat`). Il parser produrrà gli XDG, ovvero la struttura morfo-sintattica del testo in input. Seguire la seguente procedura per effettuare il parsing:
 - Avviare l'interfaccia di Chaos (`chaosgui.bat`);
 - Aprire la finestra del manager (*Fig.2-1*)
 - Avviare il server italiano (*Fig.2-2*) ed attendere che vengano caricate le basi di conoscenza (*Fig.2-3*)
 - Selezionare i file di input creati precedentemente (`[nome_frammento]_[num_paragrafo].txt`) (*Fig.2-4*)
 - Selezionare la directory di output (*Fig.2-5*)
 - Avviare il parsing (*Fig.2-6*)
 - Attendere che il parsing sia terminato (il pulsante *Fig.2-6* non indica più "please wait")
 - Nella directory di output saranno presenti i file contenenti i paragrafi parserizzati (`[nome_frammento]_[num_paragrafo].cha`)

Se si verificassero problemi durante il parsing ed i file di output non venissero prodotti, eliminare dai file di input possibili caratteri non-standard.

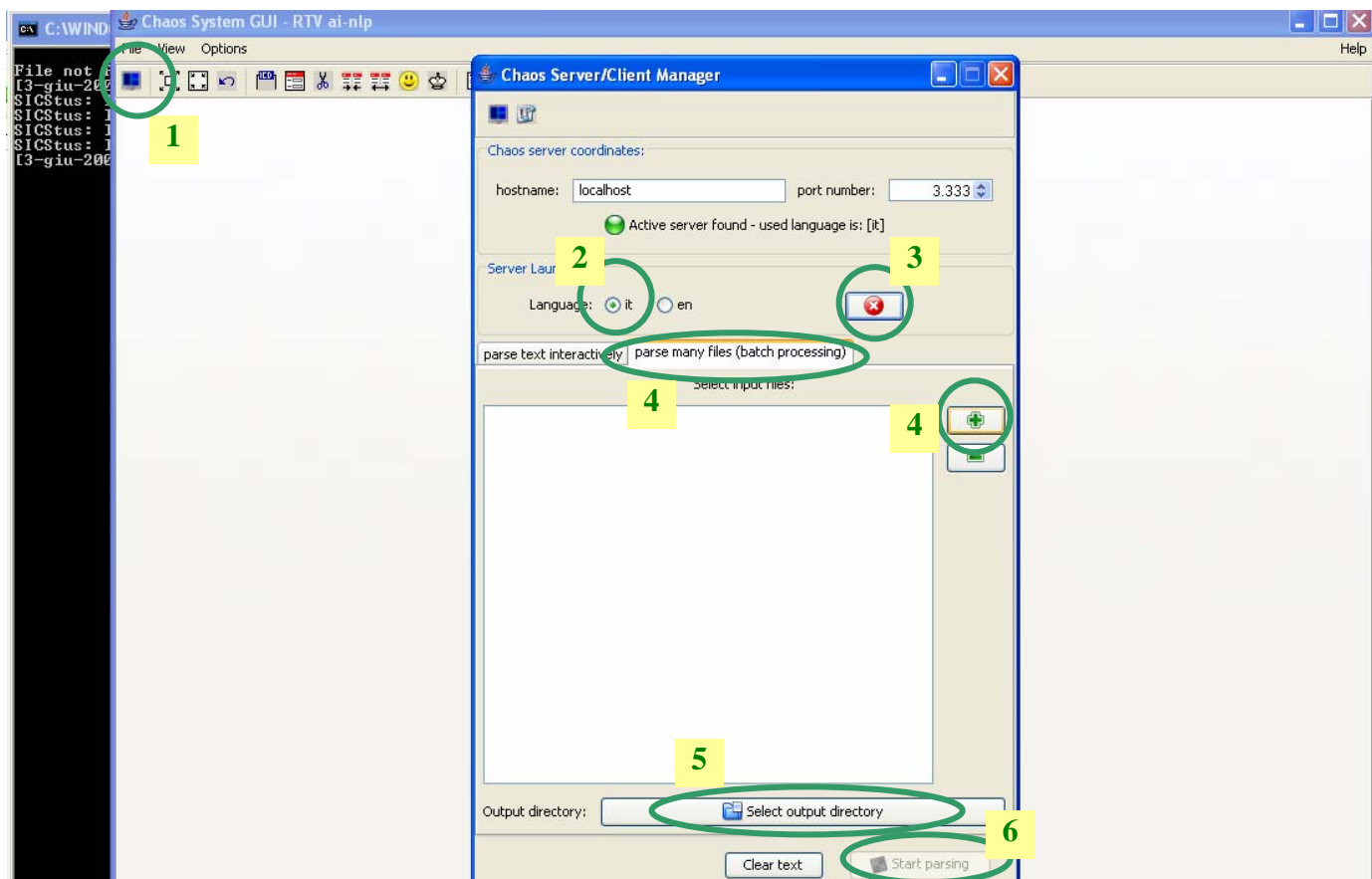


Fig 2. Come effettuare il parsing con Chaos

3. Analisi morfologica

- Per valutare l'accuracy dell'interpretazione di Chaos, utilizzare due contatori: $corr=0$, e $err=0$, indicanti rispettivamente interpretazione corretta e scorretta di Chaos su una singola parola del testo.
- Per ogni file ([nome_frammento]_[num_paragrafo].cha :
 - o Caricare il file dall'interfaccia grafica di Chaos (*File* → *Open Chaos File*)
 - o Verranno visualizzati gli XDG del paragrafo
 - o Effettuare l'analisi morfologica per ogni XDG (differenti finestre *Fig.3-1*):
 - Espandere ogni costituente (*Fig.3-2*) sino al livello di parola, utilizzando il bottone *Fig.3-3*
 - Per ogni parola aprire la *Edit Window* con il bottone *Fig.3-4*
 - Controllare se la prima interpretazione morfologica proposta da Chaos ($id:[0]$) è corretta:
 - Se è corretta: $corr+1$;
 - Se è scorretta: $err+1$;
 - Nella *Edit Window* eliminare le interpretazioni morfologiche errate (*Fig.3-5*) della parola in quel contesto frasale, utilizzando il bottone *Fig.3-6*. Lasciare una UNICA interpretazione, quella corretta. Se l'interpretazione corretta non è presente tra quelle proposte, crearla manualmente utilizzando i menù a tendina. Utilizzare la tabella delle interpretazioni morfologiche in appendice come riferimento.
 - Chiudere la *Edit Window* premendo il bottone *Fig.3-7*
 - o Terminata l'analisi di tutti gli XDG, salvare la struttura create: *File* → *SaveAs* dando al nuovo file il nome: [nome_frammento]_[num_paragrafo]_morp.cha
- Calcolare l'**accuracy** di Chaos su tutti i testi analizzati, secondo la seguente formula:

$$accuracy = \frac{corr}{corr + err}$$

4. Analisi POS tagging

- Per valutare l'accuracy dell'interpretazione di Chaos, utilizzare due contatori: $corr=0$, e $err=0$, indicanti rispettivamente interpretazione corretta e scorretta di Chaos su una singola parola del testo
- Per ogni file prodotto al punto 2 (`[nome_frammento]_[num_paragrafo]_morp.cha`):
 - o Caricare il file dall'interfaccia grafica di Chaos (*File* → *Open Chaos File*)
 - o Verranno visualizzati gli XDG del paragrafo
 - o Effettuare l'analisi di POS tagging per ogni XDG (differenti finestre *Fig.3-1*):
 - Espandere ogni costituente (*Fig.3-2*) sino al livello di parola, utilizzando il bottone *Fig.3-3*
 - Per ogni parola aprire la *Edit Window* con il bottone *Fig.3-4*
 - Nella *Edit Window* verificare se l'interpretazione di POS fornita da Chaos per la parola è corretta in quel contesto frasale. Utilizzare la tabella dei tagset in appendice come riferimento.
 - Se è corretta: $corr+1$;
 - Se è scorretta: $err+1$. Selezionare l'interpretazione corretta dal menù a tendina (*Fig.3-8*)
 - Chiudere la *Edit Window* premendo il bottone *Fig.3-7*
 - o Terminata l'analisi di tutti gli XDG, salvare la struttura create: *File* → *SaveAs* dando al nuovo file il nome: `[nome_frammento]_[num_paragrafo]_pos.cha`
- Calcolare l'**accuracy** di Chaos su tutti i paragrafi analizzati, secondo la seguente formula:

$$accuracy = \frac{corr}{corr + err}$$

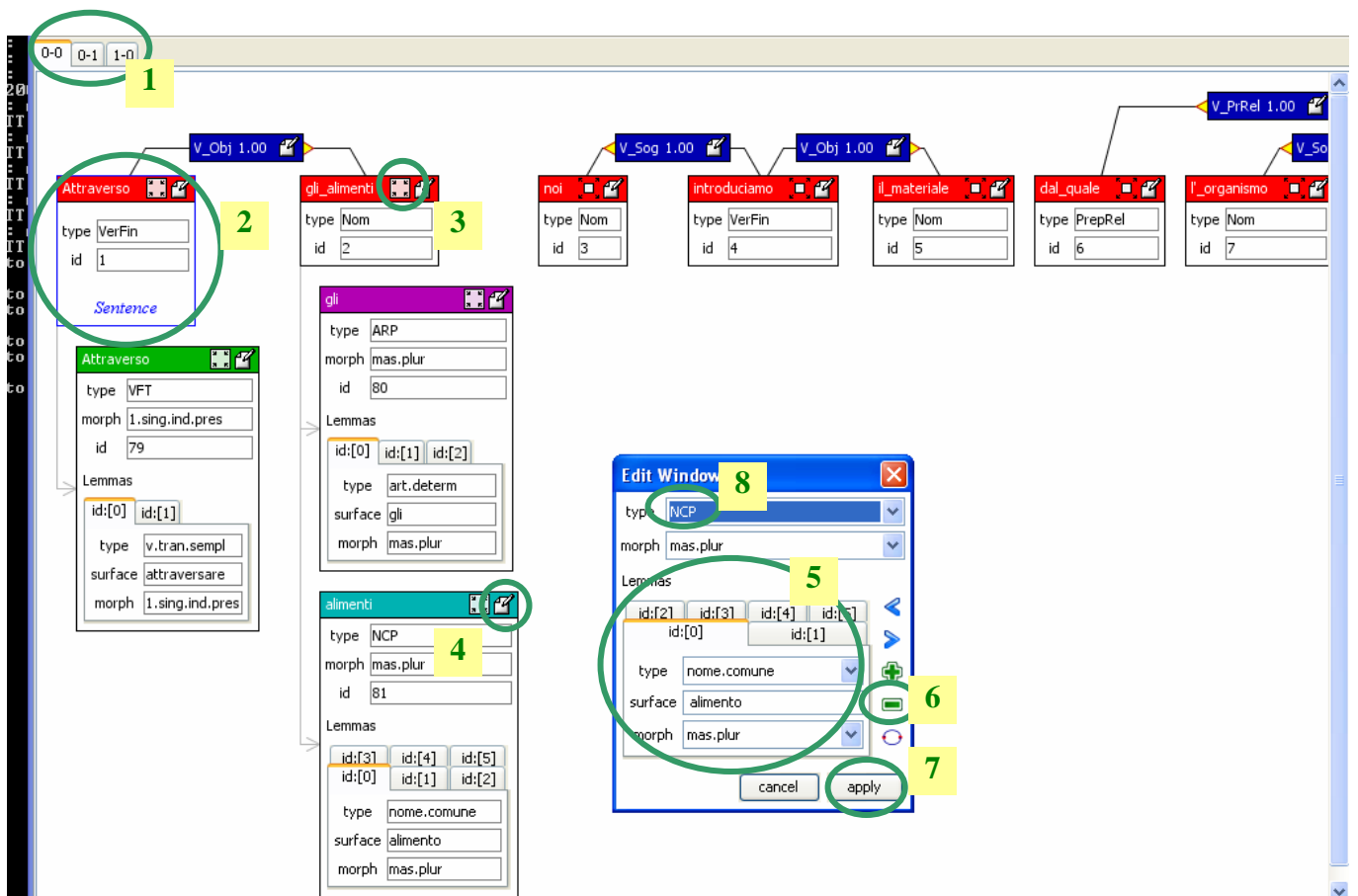


Fig 3. Come effettuare l'analisi morfologica e di POS tagging in Chaos

5. Prima di iniziare: agreement

Prima di iniziare l'annotazione complessiva, tutti i membri del gruppo dovranno annotare autonomamente (morfologia e POS tagging) il frammento di agreement `agreement_fragment.txt` scaricabile dal sito web del progetto, seguendo la stessa procedura utilizzata per l'annotazione complessiva presentata nelle precedenti sezioni.

L'annotazione finale del gruppo del frammento di agreement dovrà essere infine ottenuta confrontando le annotazioni prodotte dai singoli membri, che si accorderanno sulla interpretazioni corretta finale. Questa deve essere salvata nel file:

```
agreement_fragment_[idGruppo].cha
```

6. Invio risultati

Inviare tutti i file prodotti utilizzando la e-mail del gruppo a pennacchiotti@info.uniroma2.it.

Indicare come soggetto della e-mail: “*Progetto parte 1: gruppo [Id gruppo]*”.

Il messaggio deve contenere i valori di accuracy in questo formato:

accuracyMorp = ??

accuracyPos = ??

Allegare un file compresso (*zip* o *rar*), contenenti i file prodotti:

agreement_fragment_[idGruppo].cha

[nome_frammento].html

per ogni paragrafo:

[nome_frammento]_[num_paragrafo].txt

[nome_frammento]_[num_paragrafo].cha

[nome_frammento]_[num_paragrafo]**_morp**.cha

[nome_frammento]_[num_paragrafo]**_pos**.cha

TAGSET per l'Italiano

POS TAG	SPIEGAZIONE
AGS	Aggettivo Singolare
AGP	Aggettivo Plurale
ADS	Aggettivo Determinativo Singolare
ADP	Aggettivo Determinativo Plurale
AGI	Aggettivo Interrogativo
NUM	Numero
ARS	Articolo Singolare
ARP	Articolo Plurale
AVV	Avverbio
COA	Congiunzione Avverbiale
CPU	Congiunzione Punto
COP	Congiunzione Parentesi
COS	Congiunzione Subordinativa
CO	Congiunzione
DAT	Data
PRN	Pronome Interrogativo
PSG	Pronome Singolare
PPL	Pronome Plurale
PRR	Pronome Relativo
PR	Pronome
NCS	Nome Comune Singolare
NCP	Nome Comune Plurale
NC	Nome Comune

POS TAG	SPIEGAZIONE
NPR	Nome Proprio
PSE	Preposizione Semplice
PAS	Preposizione Articolata Singol
PAP	Preposizione Articolata Plurale
PIM	Preposizione Impropria
VX	Verbo Ausiliare
VFT	Verbo Finito Transitivo
VFI	Verbo Finito Intransitivo
VNT	Verbo NonFinito Transitivo
VNI	Verbo NonFinito Intransitivo
VNP	Verbo NonFinito TransPartPass
VIP	Verbo NonFinito IntranPartPass
VTR	Verbo Transitivo
VIN	Verbo Intransitivo
SYM	Simbolo

INTERPRETAZIONI MORFOLOGICHE PER L'ITALIANO

CLASSE CHAOS	SIGNIFICATO
Agg.dimost	Aggettivo dimostrativo
Agg.indef	Aggettivo indefinito
Agg.interr	Aggettivo interrogativo
Agg.modo	Aggettivo modale
Agg.num_card	Numero cardinale
Agg.num_ord	Numero ordinale
Agg.posses	Aggettivo possessivo
Agg.qualif	Aggettivo qualificativo
Agg.qualif.dimin	Aggettivo qualificativo diminutivo
Agg.qualif.superl.ass	Aggettivo qualificativo superlativo assoluto
Agg.qualif.superl.rel	Aggettivo qualificativo superlativo relativo
Agg.qualif_comp	Aggettivo qualificativo comparativo
Art.determ	Articolo determinativo
Art.indeter / Art.indeterm	Articolo indeterminativo
Avv	Avverbio
Avv.affermaz	Avverbio di affermazione
Avv.compar	Avverbio comparativo
Avv.dubbio	Avverbio dubitativo
Avv.giudizio	Avverbio di giudizio
Avv.interr	Avverbio interrogativo
Avv.luogo	Avverbio locativo
Avv.modo	Avverbio modale
Avv.quant / avv.quantita	Avverbio quantitativo
Avv.tempo	Avverbio temporale
Cong	Congiunzione
Cong.coord.avversat	Congiunzione coordinativa avversativa
Cong.coord.concess	Congiunzione coordinativa concessiva
Cong.coord.conclus	Congiunzione coordinativa conclusiva
Cong.coord.copul / Cong.coord.copulat	Congiunzione coordinativa copulativa
Cong.coord.dichiar	Congiunzione coordinativa dichiarativa
Cong.coord.disgiunt	Congiunzione coordinativa disgiuntiva
Cong.coord.duep	Congiunzione coordinativa
Cong.coord.esplicat	Congiunzione coordinativa esplicativa
Cong.coord.punto	Punto
Cong.coord.pvirg	Punto e virgola
Cong.coord.tempor	Congiunzione coordinativa temporale
Cong.coord.tratt	Trattino
Cong.coord.virgola	Virgola
Cong.subord.avvers	Congiunzione subordinativa avversativa
Cong.subord.caus / Cong.subord.causale	Congiunzione subordinativa causale
Cong.subord.compar	Congiunzione subordinativa comparativa
Cong.subord.concess	Congiunzione subordinativa concessiva
Cong.subord.condiz	Congiunzione subordinativa condizionale
Cong.subord.consec	Congiunzione subordinativa consecutiva
Cong.subord.dichiar	Congiunzione subordinativa dichiarativa

Cong.subord.finale	Congiunzione subordinativa finale
Cong.subord.interr	Congiunzione subordinativa interrogativa
Cong.subord.luogo	Congiunzione subordinativa luogo
Cong.subord.modo	Congiunzione subordinativa modo
Cong.subord.paren	Parentesi
Cong.subord.tempor	Congiunzione subordinativa temporale
Esclamaz	Esclamazione
Interiez	Interiezione
Nome.comune	Nome comune
Nome.comune.dimin	Nome comune diminutive
Nome.proprio	Nome proprio
Nome.straniero	Nome straniero
Prep.artic	Preposizione articolata
Prep.sempl	Preposizione semplice
Pron.dimost / Pron.dimostr	Preposizione dimostrativa
Pron.indef	Pronome indefinito
Pron.interr	Pronome interrogative
Pron.num	Pronome numerale
Pron.pers.cli.acc	Pronome personale clitico accusativo
Pron.pers.cli.dat	Pronome personale clitico dativo
Pron.pers.com	Pronome personale
Pron.pers.loc	Pronome personale locativo
Pron.pers.nom	Pronome personale nominativo
Pron.pers.par	Pronome personale
Pron.pers.rif	Pronome personale riflessivo
Pron.posses	Pronome possessive
Pron.relat	Pronome relative
Separatore	Separatore
V.intran.aus	Verbo intransitivo ausiliare
V.intran.copul / V.intrans.copul	Verbo intransitivo copulativo
V.intran.impers	Verbo intransitivo impersonale
V.intran.rifles	Verbo intransitivo riflessivo
V.intran.sempl	Verbo intransitivo semplice
v.trans._&_intran.sempl	Verbo intransitivo e transitivo semplice
v.tran.aus	Verbo transitivo ausiliare
v.tran.copul	Verbo transitivo copulativo
v.tran.rifles	Verbo transitivo riflesivo
v.tran.sempl	Verbo transitivo semplice