

Li.Di.A.

Laboratorio Virtuale Permanente di ricerca /didattica su Linguistica degli Artefatti Digitali

Stefano Penge, Andrea Sterbini, Andrea Villarini
Sapienza Università di Roma

Sommario

A. Sintesi.....	2
B. In una pagina.....	3
1. Un oggetto in cerca di analisi.....	3
2. Una professione senza professionisti.....	3
3. I risultati attesi.....	3
C. Più estesamente	4
1. Stile e differenza.....	4
1.1 Lingustica e digitale.....	4
1.2 Standard / Stile.....	4
1.3 Il punto di vista del parlante/scrivente.....	4
2. Linguistica applicata al digitale.....	4
2.1 Langue.....	4
2.2 Parole.....	5
2.3 Retorica.....	5
3. Agenda.....	6
D. Attività ipotizzabili.....	7
1. Attività di ricerca:.....	7
2. Attività didattiche:	7

A. Sintesi

Si tratta di un progetto di ricerca che incrocia tre discipline: linguistica, informatica e scienze della comunicazione.

Obiettivo del progetto è disegnare i confini e i metodi di una nuova area di ricerca, che applica i metodi e i concetti della linguistica ai linguaggi di programmazione dei computer in una visione trasversale e più ampia che li inserisce nel quadro della attività sociali di comunicazione.

B. In una pagina

Il progetto può essere descritto come la scoperta di un continente sconosciuto: quello dei codici sorgente dei programmi.

Più di 50 anni di letteratura, più di 50 lingue diverse.

Un corpus di testi dalle dimensioni quantitative enormi (l'archivio su web più noto di software OpenSource, SourceForge.net, contiene quasi un milione di "libri" diversi relativi solo agli ultimi 5 anni).

Eppure nessun'indagine, nemmeno di ricognizione, è stata condotta *finora* da un punto di vista scientifico.

1. Un oggetto in cerca di analisi

Il primo passo sarà quello della *costituzione dell'oggetto scientifico*.

Ci poniamo le domande di ogni ricercatore che scopre un nuovo territorio:

- ha le caratteristiche richieste da un oggetto "linguistico"?
- è interessante scientificamente?
- è possibile sfruttarlo dal punto di vista didattico?

2. Una professione senza professionisti

Se deve essere studiato, come oggetto scientifico, e diffuso didatticamente, il secondo passo è *mettere in campo un'équipe di ricercatori*:

- *come* può essere analizzato? con che tipo di strumenti (quantitativi, qualitativi)?
- da *chi* (con quali competenze) può essere analizzato?
- quali *risorse* sono necessarie? Persone, ambienti, macchine; ma anche relazioni ecc..

3. I risultati attesi

Il terzo passo è costituito dalla verifica delle *prime ipotesi di lavoro*:

- si possono evidenziare stili personali, o di scuola, o di periodo, o di area geografica?
- si possono istituire paralleli con le altre storie dei "testi" (con la storia della letteratura, o di un genere particolare, ma anche con la storia della musica, del teatro)?
- sulla base della sua storia, si può *prevedere* un'evoluzione?
- si può *indirizzare* un'evoluzione? si possono dare direttive sulla scrittura dei programmi?

C. Più estesamente

1. Stile e differenza

1.1 Linguistica e digitale

Sulla lingua che si parla tra umani su Internet, o su quella che si dovrebbe parlare (web writing ecc.) esiste ormai una letteratura di rispetto, mentre sulla conversazione tra essere umano e essere digitale ancora stato scritto troppo poco. Alcune aree periferiche (net-art, letteratura combinatoria, code-work, Perl poetry, ...) sono state appena esplorate, ma mai il centro: la/le lingue con cui, da oltre 50 anni, in tutto il mondo, si scrivono i programmi per computer.

1.2 Standard / Stile

Probabilmente questo oggetto è sempre sfuggito all'analisi perché si è diffusa la convinzione che programmare sia un'attività logica, oggettiva, matematica, e che consista semplicemente nel trovare la forma giusta, l'algoritmo più adatto allo scopo. Se l'efficienza è presa come unica guida, e lo standard come modello, ogni deviazione va perseguita e non studiata (o studiata con fini "terapeutici"). Deviazione come devianza.

Questo è il punto di vista dell'azienda, e in parte dell'accademia, che cerca di insegnare il buono stile di programmazione (da Wirth a Knuth) giustificandolo in termini di diminuzione di costi globali (non più solo efficienza di esecuzione, ma abbassamento dei costi di archiviazione, manutenzione, estensione, personalizzazione).

1.3 Il punto di vista del parlante/scrivente

Ma, sorpresa: il punto di vista dei programmatori è diverso (quando ci si prenda la briga di intervistarli): per loro programmare è un'arte, a volte a prescindere dai risultati. La qualità è l'unica vera guida, e l'efficienza è solo una delle sue declinazioni. Uno stesso problema può essere risolto in tanti modi diversi quanti sono i programmatori che lo affrontano. Se si ascoltano i programmatori, l'aspetto personale, stilistico, cioè la variazione emerge sempre. Qui "stile" significa proprio la stessa cosa che "stile poetico": qualcosa di unico e di percepibile, che identifica i lavori di un autore.

2. Linguistica applicata al digitale

2.1 Langue

La linguistica di questo linguaggio artificiale deve ancora nascere. Esistono ovviamente grammatiche specifiche dei singoli linguaggi (non grammatiche descrittive, ma prescrittive); non ci sono però studi che trattano i linguaggi di

programmazione diversi allo stesso modo delle lingue naturali, con caratteristiche espressive, specificità, evoluzione. Al massimo si tentano derivazioni di un linguaggio dall'altro, o se ne misura la diffusione nel tempo e nello spazio, senza arrivare ad una linguistica degna di questo nome. In ogni caso, si resta confinati nella linguistica della langue, come se il concetto stesso di parole non avesse senso, non potendo in questo dominio la lingua in atto discostarsi significativamente da quella in potenza.

Nessuna delle "scoperte" della linguistica moderna (dalla stilistica alla semiotica del testo alla linguistica attanziale o acquisizionale) è mai stata applicata a questo campo.

2.2 Parole

Superando questa barriera, e guardando i codici dei programmi come testi, si scoprono cose interessanti. Il parallelismo, ad esempio, tra scrittura di programmi e la scrittura di testi poetici è forte:

- un linguaggio fortemente controllato, elitario (l'apologia della poetica del constraint);
- testi chiusi, che mirano a produrre effetti e non a veicolare informazioni;
- autori socialmente "inadatti" che lavorano nell'ombra (il mito dell'hacker);
- circoli iniziatici, scuole di pensiero;

Naturalmente il destinatario cambia (ma fino a che punto? anche i programmi, come le poesie, vengono letti e giudicati da umani, oltre che da altri programmi).

E si possono immaginare altri parallelismi: per esempio con la scrittura musicale. Anche lì il testo non è scritto per essere letto, ma eseguito; e tuttavia si possono immaginare letture non esecutive di testi musicali. Ci sono testi scritti per essere guardati (Satie) prima che ascoltati.

Stili, scuole, mode: si può tracciare un'intera storia della programmazione da questo punto di vista. Ogni programmatore, in quanto autore nella pienezza di questo termine, ha gusti e idiosincrasie personali, individuabili e forse addirittura riscontrabili trasversalmente su tutti i tipi di testi che produce (nelle lingue naturali e non).

2.3 Retorica

Se la linguistica dell'artificiale è lontana dall'essere disciplina affermata e riconosciuta, non è stata ancora nemmeno tentata un'analisi *retorica* dei programmi, in quanto testi scritti con lo scopo di convincere (una macchina, ma anche altri programmatori/manutentori). In fondo la prospettiva poetica non rende conto appieno del ruolo principalmente funzionale dei testi artificiali, mentre quella retorica offre 2500 anni di riflessioni e indicazioni didattiche sul discorso che muove. Si possono allora adattare e cercare figure e stilemi e studiarne il ruolo, il rapporto. Qui il paragone con il testo pubblicitario sembrerebbe poter aiutare.

3. Agenda

Si tratta di acquisire un corpus di testi, di intervistare autori, di cominciare a delineare ipotesi, effettuare verifiche.

Esiste ormai un corpus di testi molto esteso (grazie anche al paradigma di sviluppo OpenSource, che ha avuto tra i suoi effetti l'archiviazione dei testi dei programmi per permetterne la modifica) che permetterebbe di effettuare persino analisi di tipo statistico con strumenti quantitativi.

Dovranno essere coinvolte aziende che hanno ospitato e conservato l'opera di autori di programmi (anche come sponsor). Si dovrà creare un laboratorio permanente, anche se solo virtuale, per permettere scambio di testi, di ipotesi, di strumenti. Si dovrà creare un *équipe* di lavoro mista, interdisciplinare, internazionale, dotata di forte spirito di collaborazione e di grande capacità di vedere analogie.

Sarà forte la ricaduta didattica, intesa tanto come apertura di un orizzonte completamente nuovo per gli studenti delle scienze della comunicazione e delle scienze del linguaggio, quanto come un aumento di consapevolezza linguistica e "stilistica" per gli studenti delle scienze dell'informazione. Impatti immediati sono prevedibili, ad esempio, sulla valutazione delle competenze di programmazione di uno studente sulla base di un'analisi stilistica dei suoi "testi", comparata a modelli esistenti, o a serie storiche. Si potrebbero anche ipotizzare delle ricadute nel campo della didattica/acquisizione delle lingue qualora si scoprissero degli universali nell'apprendimento dei linguaggi di programmazione.

D. Attività ipotizzabili

1. Attività di ricerca:

- apertura di un sito web di lavoro /collaborazione (repertorio di testi, forum di discussione, ...): <http://twiki.di.uniroma1.it/twiki/view/Lidia>
- creazione di un laboratorio virtuale permanente inter-facoltà
- acquisizione dati e catalogazione
- definizione di ipotesi di lavoro
- creazione di strumenti di indagine quantitativa
- analisi del corpus di testi/programmi open-source
- ...

2. Attività didattiche:

- attivazione di seminari
- chiamate di ricerca
- assegnazione di tesi di laurea triennale o magistrale /dottorato
- più avanti, attivazione di insegnamenti all'interno di corsi di laurea, master, corsi di alta formazione, ...

3. Attività di promozione:

- articoli
- organizzazione di un convegno internazionale di presentazione
- partecipazione a convegni su temi affini
- pubblicazione di un testo a stampa in almeno due lingue