SCUOLA
NORMALE
SUPERIORE
PISA

Chiara Celata & Pier Marco Bertinetto

Per un'analisi morfologica del lessico italiano

(versione estesa della comunicazione per "Filologia, Linguistica e Corpora", Roma, 18 settembre 2010)

1 Introduzione

In questo contributo si illustrano i primi risultati di un progetto di ricerca incentrato sull'analisi morfologica del lessico italiano, con particolare riferimento alla morfologia derivativa e alla formazione di parole prefissate e suffissate.[1]

La ricerca, tuttora in corso di svolgimento, è volta all'analisi di alcune variabili di natura morfologica presenti nel lessico italiano, per come esso è rappresentato nel *Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS)* (Bertinetto et al. 2005). La ricerca si pone un duplice scopo. Da un lato, la meta principale è quella di arricchire il nucleo delle informazioni contenute nel corpus, specificamente nella sottosezione denominata 'Lemmario', corredando le indicazioni di natura grammaticale e statistica già presenti con una serie di informazioni sullo status morfologico delle forme (e, in futuro, sulle regolarità sillabiche e accentuali) ivi rappresentate. Dall'altro, la ricerca vuole procedere anche con verifiche *in itinere* della validità e dell'adeguatezza delle operazioni di codifica predisposte per l'inserimento delle informazioni nuove, estraendo cioè campioni rappresentativi di dati di tipo prevalentemente morfologico (tipi di affissi, radici, allomorfi etc.), analizzandone le caratteristiche quantitative (ricorrenze assolute e relative etc.), e cercando un riscontro diretto con quanto già parzialmente contenuto (ma non sempre empiricamente verificato) nella bibliografia specialistica su classi e regole morfologiche dell'italiano.

[1] Il progetto, che comprende anche l'analisi della morfologia flessiva, nonché una parte relativa all'analisi prosodica (sillabica e accentuale) del lessico italiano (di cui non ci occuperemo in questo contributo), viene svolto nell'ambito del finanziamento quandriennale FIRB 2009-2012 "WIKIMEMO.IT: Il portale della lingua e della cultura italiana", di cui P.M. Bertinetto è coordinatore nazionale.

2 Il CoLFIS

Il CoLFIS comprende quasi quattro milioni di parole (per la precisione: 3.798.275 ricorrenze lessicali, tratte da quotidiani del periodo 1992-1994, periodici e libri), ricavate da uno scrupoloso campionamento di testi scritti scelti sulla base di dati ISTAT 1993 relativi alle letture preferite dagli italiani.

Il corpus nasce esplicitamente con lo scopo di essere il più possibile rappresentativo del lessico mentale di un parlante italiano di media cultura (Laudanna et al. 1995: 105-106). Le finalità e gli interessi, infatti, che muovono il progetto di costituzione del CoLFIS sono, fin dall'inizio, di tipo psicolinguistico:

> "Quando in un contesto psicolinguistico si parla della frequenza di parola, ci si riferisce ad una ideale frequenza media, che faccia astrazione da particolari idiosincrasie, frequenze soggettive, sovraesposizioni di talune categorie di persone a lessici specialistici, etc. E' perciò necessario che un lessico di frequenza centrato sul ricevente non sia basato su tipi testuali stabiliti a priori o campionati in maniera distorta, ma su criteri di selezione del corpus che permettano di far emergere per quanto possibile la frequenza media di ricezione." (Laudanna et al. 1995: 104).

Per corpus 'dalla parte del ricevente' si intende quindi un corpus che rappresenti nel modo più fedele possibile il tipo di lessico effettivamente letto (piuttosto che quello più frequentemente prodotto), quindi ricavato da una scrupolosa campionatura dei testi maggiormente diffusi nelle letture degli italiani, e rappresentati proporzionalmente all'ampiezza della loro fruizione. In questo senso, il CoLFIS si differenzia notevolmente da altri dizionari di frequenza dell'italiano scritto in uso (es. LIF, Bortolini et al. 1971), non solo per il fatto di includere un numero decisamente superiore di lemmi (quasi quattro milioni di voci, a fronte delle circa 500.000 parole del LIF), ma anche e soprattutto per il bilanciamento delle fonti, che mira a rispecchiare la frequenza effettiva dei diversi usi scritti della lingua.

Il corpus nella su aveste attuale si articola fondamentalmente nelle due grosse sezioni 'Formario' (che contiene la frequenza di ciascuna forma presente nel corpus) e 'Lemmario' (che contiene la frequenza delle forme lemmatizzate, incluse le parole 'sintagmatiche'), nonché una terza sezione 'ibrida' ('Lemmi e forme'). Al loro interno, le sezioni sono strutturate in diversi tipi di organizzazione dei dati (rispetto ai

parametri dell'ordine – diretto, inverso, numerico – e del carattere – maiuscolo, minuscolo, numerico –).

Date queste caratteristiche, il corpus si presenta dunque come uno strumento unico di arricchimento delle conoscenze sul lessico italiano (e sulle categorie grammaticali ivi rappresentate), in particolare per quanto riguarda gli usi psicolinguistici di tali conoscenze. Al tempo stesso, il patrimonio lessicale rappresentato dal CoLFIS presenta potenzialità enormi per quanto concerne, oltre che il lessico, anche le sue sotto-strutture. Si pensi infatti alla possibilità, a partire da una banca-dati lessicale di queste dimensioni, di derivare informazioni quantitative non impressionistiche relativamente a morfemi di livello inferiore alla parola (e in particolare, a morfemi flessivi e derivazionali). E' noto infatti che non solo la frequenza lessicale, ma anche la frequenza dei costituenti morfologici, siano essi radice o affisso, ha effetti chiari e potenti sull'elaborazione delle parole da parte dei parlanti. Le potenzialità del CoLFIS rispetto agli aspetti empirici di questo ambito di studi sono a tutt'oggi ancora quasi completamente inesplorate. D'altra parte, ricerche psicolinguistiche su riconoscimento, comprensione, produzione o lettura di parole morfologicamente complesse sono (forse fin dagli inizi di questo settore di studio) numerose e feconde per quanto riguarda l'italiano (per una sintesi aggiornata, vedi Burani 2006).

Qualcosa di simile può valere anche per gli ambiti dell'elaborazione fonetico-fonologica e (per quanto riguarda il linguaggio scritto) ortografica. Menzionavamo sopra, infatti, che tra gli obiettivi allargati del presente progetto di ricerca sul CoLFIS rientra anche la possibilità di corredare la banca-dati attualmente disponibile con informazioni circa gli elementi fonologico-prosodici rispetto a cui si differenzia e struttura il lessico italiano.

Il presente lavoro si inserisce dunque, in misura pienamente coerente con le ragioni che per prime hanno mosso alla costituzione del CoLFIS, dentro ad un complesso progetto di trasformazione dell'attuale lessico di frequenza in uno strumento per indagini quantitative su molteplici aspetti di ordine sub-lessicali.

3 Descrizione del progetto sulla morfologia derivazionale

La morfologia derivazionale rappresenta un comparto particolarmente importante dell'organizzazione morfologica dell'italiano. Secondo il *Grande Dizionario Italiano dell'Uso* (UTET, 2003, curato da T. De Mauro), conta circa 93.000 derivati, oltre un terzo del totale dei lemmi catalogati. Vengono identificati 91 prefissi (distribuiti su

circa 17.000 lemmi) e 316 suffissi, divisi in suffissi verbali, avverbiali, nominali e aggettivali (questi ultimi due di gran lunga più numerosi degli altri), che danno origine alla grande maggioranza dei derivati italiani (è ben nota la cosiddetta preferenza 'universale' per la suffissazione, chiaramente rappresentata in lingue come l'italiano).

L'attuale progetto di ricerca su CoLFIS è finalizzato ad introdurre, per ogni forma del Lemmario, informazioni circa la presenza di morfemi derivativi (prefissi o suffissi), e quindi, nel caso delle forme suffissate e prefissate, circa gli aspetti formali relativi alla base e gli aspetti semantici relativi al processo di formazione della parola.

Da un punto di vista estremamente concreto, ciò che si vuole ottenere è una codifica che permetta, al ricercatore interessato agli aspetti quantitativi della prefissazione e suffissazione italiana, di differenziare un affisso derivazionale (es. -*bile* in *leggibile*, -*mente* in *velocemente*) da una sequenza omografa (es. -*bile* in *flebile*, -*mente* in *deprimente*). Allo stato attuale, infatti, il CoLFIS non permette di raccogliere informazioni quantitative relativamente a morfemi di livello inferiore alla parola, ma solo ed esclusivamente di elencare le forme che 'iniziano' o 'finiscono' con una certa sequenza di grafemi (o, anche, che la contengono al loro interno).

Le procedure di base e gli aspetti problematici finora riscontrati nel corso delle operazioni verranno illustrati con maggiori dettagli nel prossimo paragrafo. È necessario però puntualizzare fin d'ora che tra i prodotti che si vogliono ottenere non rientra la *segmentazione* delle forme, ma esclusivamente la segnalazione della presenza, al loro interno, di eventuali morfemi formativi.

Le operazioni non si svolgono nel rispetto dell'ordine delle entrate lessicali (ordine 'alfabetico' complessivo), ma per sottoinsiemi di lessico definiti sulla base della presenza di determinati affissi. Gli affissi sono stati scelti e catalogati in accordo con l' 'ipotesi della base unica' (Scalise 1994: 210-217), secondo cui un affisso si aggiunge solo a basi che formano una classe sintattica definibile come [+N] o come [+V]. È in corso l'analisi e la codifica delle forme composte con alcuni dei prefissi e suffissi più produttivi in italiano, e in particolare: -aceo/-accio, -aglia/-aia, -aio1(locativo)/-aio2(agente)/-ario/-aro, -bile/-evole, -crazia/-crate/-cratico, -ente/-enza/-ante/-anza, -ezza/-izia/-igia, -iere/-iera/-iero, -ino1(diminutivo) e -ino2(etnonimo), -etto, -ismo/-ista/, -logia/-logo, -mente, -mento, -tore, -zione (per i suffissi), auto-, bi-/bis-, de-/di-, in1(direzionale)- e in2(negativo)-, mini-/micro-, re-/ri-, s1(contrario)-/s2(iterativo)/dis-, tra-/tras-/trans- (per i prefissi).

Il progetto procede in maniera incrementale all'interno della morfologia derivazionale (aggiunta di nuovi affissi, codifica di forme in cui sono presenti più

affissi alla volta etc.). Come anticipato prima, in una fase successiva si provvederà all'introduzione della codifica di alcune categorie fondamentali della flessione verbale e nominale, e infine si tenterà l'introduzione, stavolta automatica, di informazioni di natura sillabica e accentuale.

4 Criteri di analisi, problemi empirici e teorici

Se è vero che le parole si dividono in semplici (es. *ieri*) e complesse (es. *utilità*) (Scalise 1995: 473), è anche vero che nel lessico (italiano, come di molte altre lingue) si incontrano piuttosto comunemente forme dallo statuto morfologico incerto. Le ragioni che hanno dato origine a forme dubbie sul piano dell'analisi morfologica possono essere le più disparate: ad esempio, può succedere che quello che in diacronia si ricostruisce chiaramente come derivato non lo sia più, o non più così chiaramente, in sincronia, nella percezione del parlante (es. *mobile*, *flebile*); oppure, al contrario, che una forma mantenga intatto il significato morfologico dell'affisso, ragione per cui il parlante tende a classificarla come forma derivata, pur non essendo riconducibile a nessuna forma base in sincronia (es. *protagonista*). Vari possono dunque essere i processi e le condizioni che portano il giudizio dei parlanti (più o meno consciamente espresso) a vacillare rispetto al grado di complessità morfologica delle forme; tale incertezza è il prodotto dell'azione, spesso incrociata, di fattori sia formali, sia semantico-interpretativi.

La riconoscibilità delle sottoparti di una parola morfologicamente complessa si misura in termini di 'trasparenza': una parola morfologicamente complessa potrà essere più o meno trasparente (e quindi, più o meno opaca) rispetto alla sua base, quanto più riconoscibili (o irriconoscibili) saranno i componenti che la costituiscono. È chiaro quindi che la trasparenza è una proprietà scalare, non categorica: si potrà ad esempio dire che il rapporto tra *unzione* e *ungere* è meno trasparente di quello tra *osservazione* e *osservare*. Sia *unzione* che *osservazione* sono, comunque, parole chiaramente derivate. Vi sono casi, però, in cui il rapporto tra base e derivato può essere (o diventare) così poco trasparente, da dar luogo a interpretazioni per cui il concetto stesso di parola morfologicamente complessa sfuma irrimediabilmente verso il polo della non-complessità. Di alcuni di questi casi parleremo qui sotto, a testimonianza del fatto che l'aggiunta di informazioni sulle strutture sub-lessicali nel CoLFIS costituisce un'operazione delicata sul piano operativo e ricca di spunti importanti sul piano teorico.

La scalarità del concetto di trasparenza morfologica spiega quindi bene perché, analizzando i derivati italiani, si abbia talvolta l'impressione di trovarsi davanti ad un *continuum* di possibilità della derivazione, ai cui estremi stanno, rispettivamente, piena trasparenza e cristallizzazione della struttura derivazionale delle parole. Un *continuum* certamente non simmetrico, per il fatto di presentare un addensamento (in termini quantitativi) delle forme intorno al polo della trasparenza (anche se questa può essere ridotta o oscurata sotto l'effetto di determinati fattori formali e semantici); ma rimane pur sempre evidente l'esistenza di un bacino di forme che per il parlante nativo (non linguista) possiedono statuto morfologico incerto.

I rapporti tra base e derivato si misurano in termini formali e semantico-interpretativi. Nel primo caso (trasparenza/opacità morfotattica; cf. ad esempio Gaeta 2002 per un approccio naturalista), la relazione tra le due forme viene opacizzata dall'intervento di regole di riaggiustamento morfo-fonologico che introducono una variabilità morfologicamente immotivata, come nel caso di *balcone* – *balconcino* (**balconino*) rispetto a *tavolo* – *tavolino*. Nel secondo caso, i rapporti semantici originari tra il derivato e la sua base non sono più disponibili al parlante, che non istituisce più nessuna connessione di tipo semantico tra il derivato e la base, pur riconoscendo il primo come forma morfologicamente complessa. Un esempio in tal senso può essere rappresentato dall'italiano *orazione*, che i parlanti riconoscono correttamente come un derivato in *–zione*, ma la cui base è semanticamente opaca; oppure, si possono citare casi come quello di *protagonista*, che si classifica come appartenente al gruppo delle parole derivate in *–ista* (il suffisso è formalmente integro e non oscurato da regole di riaggiustamento fonologico), anche per via della 'solidarietà' con la formazione corrispondente in *–ismo* (*protagonismo*, cf. *comunista* – *comunismo*), nonché per effetto dell'esistenza di forme come *agonista* (e *agonismo*), *antagonista* (e *antagonismo*) etc., ma rispetto al quale la base nominale è irrecuperabile (almeno, per un parlante di cultura non classicistica).

A rigore, i casi di opacità morfo-fonologica del tipo di quella menzionata sopra (v. *balconcino*) non rappresentano di per sé casi particolarmente problematici per quanto riguarda le operazioni di codifica attualmente svolte sul CoLFIS, tanto più che, come già detto, il lavoro nella sua fase attuale non prevede la segmentazione delle forme; le allomorfie, sia delle basi, che degli affissi, non richiedono *a priori* una elaborazione specifica (ciò vale anche per i casi estremi di suppletivismo della radice).

Ad una interazione tra opacizzazioni morfo-fonologiche e particolarità semantiche nel rapporto tra base e derivato vanno invece attribuiti i casi di non facile

riconducibilità del derivato rispetto alla sua forma base. Questi aspetti si sono rivelati particolarmente problematici per l'analisi dei dati italiani e hanno richiesto un livello di elaborazione maggiore per quanto riguarda la codifica delle informazioni morfologiche. Gli esempi che seguono sono infatti mirati ad esemplificare le questioni relative al trattamento di quelle forme derivate che abbiamo giudicato essere particolarmente distanti dal 'prototipo' (se esiste) di derivato in italiano.

Abbiamo visto che allomorfie anche non produttive, ma chiaramente analizzabili, non presentano problemi specifici di codifica; una forma come *fattibile* (rispetto a **facibile*, come richiederebbe la regola di formazione per cui *-bile* si aggiunge al tema verbale) è inequivocabilmente un derivato come *amabile*. Problemi invece nascono per quei casi in cui il suppletivismo della base è così estremo, che la base non è riconducibile a nessuna forma attestata in sincronia (es. *potabile*, *rettore*). Da questo punto di vista, casi come *protagonista* visto sopra e casi come *potabile* o *rettore* sono alla fine dei conti molto simili, per il fatto di presentare un rapporto semantico-etimologico con la base estremamente indebolito. In sincronia, cioè nella percezione del parlante, forme di questo tipo sono o non sono derivati? Parrebbe proprio di sì, visto che l'integrità del suffisso (presente e riconoscibile in esse) non è solo formale, ma anche semanticamente ricco di significato (*potabile* è effettivamente "qualcosa che può essere V + part. pass", salvo recuperare la radice verbale e il suo significato; *rettore* è certamente un "agente che compie l'azione di V", salvo recuperare il senso di questo V). Si tratta pertanto di forme non solo diverse, sul piano della struttura morfologica, rispetto a forme omografe non complesse come *vista* o *motore*, ma anche diverse rispetto a *flebile* o *abile*, e vanno quindi classificate appropriatamente.

Per rendere conto di fenomeni alla 'periferia' delle regole di formazione di parola come quelli appena citati, è stato deciso di introdurre nella codifica, accanto all'informazione sulla trasparenza semantica tra base e derivato, anche una informazione sul grado di autonomia morfosintattica della presunta 'base' (cioè un criterio per stabilire se essa rappresenti un morfema in un certo senso 'legato', dentro alla parola derivata). Esemplificheremo con alcuni esempi tratti dal gruppo dei derivati in *–bile*, un suffisso che presenta diverse particolarità relativamente alla tipologia di combinazioni attestate (Grossmann & Rainer 2004).

La base di una parola derivata può essere non riconducibile ad una forma base in sincronia (cioè, semplificando, presentarsi come una radice 'legata') per diversi motivi, e in diversi gradi. Nella forma più semplice, essa può essere fonologicamente non coincidente con la base che la specifica regola di formazione di parola

richiederebbe, ma al contempo essere facilmente recuperabile. È, ad esempio, il caso di parole come *visibile* o *comprensibile* (costruite a partire da una forma di participio perfetto latino italianizzato). Ad un livello leggermente superiore di problematicità, si situano quei derivati la cui base, oltre ad essere fonologicamente eccentrica, è anche difficilmente recuperabile, pur mantenendosi la possibilità (almeno, per un parlante mediamente istruito) di recuperarne in senso generico il significato o la sfera semantica di appartenenza. Si tratterebbe dei casi come *compatibile* (nel senso di "adattabile", "conciliabile" con qualcosa) o *vulnerabile*. Infine, il caso più marcato di subordinazione morfosintattica della base è rappresentato da quei derivati le cui basi sono individuabili solo in termini di sfera semantica generica di appartenenza, e, in più, i cui affissi hanno perso (o mantengono in misura estremamente labile) il 'significato della derivazione' (o *Wortbildungsbedeutung*, Rainer 2004:22) e, con esso, la funzione semantico-sintattica originaria: in questo gruppo faremmo rientrare termini come *responsabile*, *possibile*, *terribile*, *sensibile* (ad esempio, nella semantica di *possibile* rientra qualcosa che ha a che fare con il 'potere' o la 'possibilità', ma *possibile* non è più parafrasabile in senso stretto come "che può essere V + part. passato"; qualcosa di molto simile può valere di volta in volta per gli altri esempi citati). A rigore, una forma come *potabile* è ancora diversa (forse speculare) rispetto a *possibile* e simili, per il fatto – come abbiamo già detto – di presentare integro il valore semantico-sintattico dell'affisso, e contestualmente di aver perso la recuperabilità della base. In un certo senso, potremmo dire che nei derivati come *potabile* la radice è recuperabile 'per sottrazione', cioè 'in negativo' rispetto all'affisso, mentre in *possibile* la radice è recuperabile 'in positivo' limitatamente al riferimento formale (in termini, cioè, fonologici) ad una certa classe di parole specificate per una semantica comune.

La casistica qui elaborata potrebbe dunque rappresentare un buon esempio di quella 'scalarità' empiricamente attestata per i fenomeni di derivazione in italiano, cui accennavamo sopra. Potremmo dunque utilizzare uno schema come in Tabella 1 per rappresentare la gradualità del passaggio da una forma 'prototipicamente' derivata, come *amabile*, ad una ormai non più classificabile come tale, come *nobile*, di cui le grammatiche predicano la non complessità morfologica. I parametri discriminanti sono quello della trasparenza semantica del rapporto tra base e derivato, e quello dell'autonomia morfosintattica della base, entrambi realizzati tramite opzione binaria. Per mezzo di questi strumenti, abbiamo tentato di ricondurre le disparità di

comportamento dei diversi derivati italiani ad una griglia di informazioni facilmente recuperabili in associazione ad ogni singola entrata lessicale.

	esempi	TRASPARENZA SEMANTICA	AUTONOMIA MORFOSINTATTICA DELLA BASE
1.	*amabile*	+	+
2.	*visibile* *comprensibile*	+	+
3.	*compatibile*[2] *vulnerabile*	+	-
4a.	*responsabile* *possibile terribile* *sensibile*	-	-
4b.	*potabile*	-	-
5.	*nobile* *flebile* *abile*	parole non derivate	

Tabella 1. Esempi di composti in –bile e omografi non derivati

Si noteranno in particolare due cose.

In primo luogo, il gruppo 1 e il gruppo 2 non si differenziano né sulla base del criterio semantico, né sulla base del criterio morfosintattico. Ciò corrisponde al fatto che, sulla base dello spoglio dei dati finora realizzato, si è deciso che le parole del gruppo 2, pur essendo caratterizzate da un certo grado di deviazione rispetto alla regola generale di formazione delle parole in *–bile*, non debbano essere classificate in modo diverso rispetto alle parole del gruppo 1.

Il secondo aspetto da mettere in evidenza, in una classificazione come quella di Tabella 1, riguarda la criticità della distinzione tra il gruppo 3 e il gruppo 4b, distinti solo per il presunto livello di trasparenza semantica del rapporto tra base e derivato (si

2 Come specificato sopra, si fa riferimento qui a *compatibile* esclusivamente nel senso di "adattabile", "conciliabile" con qualcosa (e non nel senso di "che può essere compatito").

ricorderà che il gruppo 4a è parzialmente diverso, per il fatto che per opacità semantica si deve specificamente intendere la perdita del valore semantico e funzionale dell'affisso, piuttosto che della base). La recuperabilità semantica di una base rispetto ad un derivato può essere soggettiva, e non è chiaramente quantificabile quanto la semantica di basi formalmente inesistenti possa conservarsi nel derivato. Questo esempio, d'altra parte, illustra bene come il progetto che stiamo descrivendo sia per sua stessa natura (cioè, per il fatto di servirsi di un incrocio di parametri indipendenti per esprimere lo status morfologico delle forme) esposto ad una tensione continua tra il rischio di iper-caratterizzare e quello di ipo-caratterizzare le entrate rispetto alle informazioni morfologiche rilevanti.

Non solo quando la base è irrecuperabile (o difficilmente recuperabile), ma anche con i parasintetici si hanno chiari esempi di forme derivate il cui status è in qualche modo diverso da quello delle forme derivate prototipiche. Per parasintetici si intendono parole complesse formate da almeno tre elementi, cioè una base e due o più morfemi legati che vengono aggiunti simultaneamente a destra e a sinistra della base, come ad esempio *sfacciato* (Scalise 1995: 511). Parole di questo tipo presentano un alto grado di trasparenza semantica (la base è *faccia*), ma la radice non è pienamente autonoma come nei derivati non parasintetici formati da un processo di prefissazione più un processo di derivazione (es. *invariabilmente*): alla base *faccia* sono stati aggiunti il prefisso *s-* e il suffisso *–ato*, ma non esistono né **facciato* né **sfaccia*. Sulla base dei parametri di analisi individuati sopra, la codifica delle forme di questo genere sarà quella di derivati caratterizzati da [+ trasparenza semantica] e [– autonomia morfosintattica della base], intendendo però in questo particolare caso la base come, di volta in volta, **facciato* oppure **sfaccia* (e non quindi *faccia*). In un parasintetico derivato da *–bile* come *impensabile*, si ha un chiaro esempio di un fenomeno che riveste un'importanza non marginale ai nostri scopi, e cioè l'effetto legittimante di un affisso rispetto alla derivazione tramite altro affisso: **impensare* non esiste e (*)*pensabile* è considerato una retroformazione (*in-* ha un effetto legittimante rispetto a *–bile*).

Infine, menzioneremo un ultimo problema attinente alle classificazioni su cui si reggono le operazioni di codifica, relativo ad un'altra dimensione di variabilità dei derivati: specificamente, alla possibilità che una medesima forma vada collocata su gradini diversi di quella scala di 'derivazionalità' ipotizzata sopra (ed esemplificata con alcuni fenomeni relativi al suffisso *–bile*, vedi Tabella 1), a seconda della categoria grammaticale in cui si realizza. In altre parole, ci chiediamo se, nella

percezione del parlante nativo, una parola possa essere giudicata come derivata per affissazione oppure non derivata (cioè semplice) a seconda che l'uso che se ne fa sia un uso nominale o aggettivale eccetera. Due esempi, sempre dal comparto della derivazione in *–bile*, possono esemplificare quanto detto: *mobile* e *stabile*. In quanto sostantivi (*mobile* 'oggetto d'arredo' e *stabile* 'edificio'), è plausibile che non conservino nulla, in sincronia, della semantica etimologica, e il rapporto morfo-fonologico e morfosintattico tra *mo-* e *muovere* pare estremamente opacizzato (lo stesso non può dirsi di *sta-* e *stare*, d'altra parte). Classificheremmo quindi *mobile* (s.) e *stabile* (s.) come parole non derivate (alla stregua di *nobile* e *flebile*). Nel caso di *mobile* (a.) e *stabile* (a.), però, la semantica dell'affisso (se non anche quella della base, in termini generici) è sicuramente recuperabile (un oggetto è 'mobile' quando è "qualcosa che può essere mosso", è 'stabile' quando è "qualcosa che può essere fatto stare (in qualche posto)"); in questo senso, i due termini si avvicinerebbero al caso di *compatibile* ([+ trasparenza semantica] e [– autonomia morfosintattica della base]) oppure anche a quello di visibile ([+ trasparenza semantica] e [+ autonomia morfosintattica della base]).

5 Conclusioni

Abbiamo visto che le operazioni di codifica delle informazioni morfologiche relativamente ai fenomeni di derivazione in italiano sul CoLFIS presentano ricadute sul piano teorico per quanto riguarda l'elaborazione di concetti chiave di questo ambito di studi, in particolare per il fatto di obbligare il ricercatore a rendere conto di una scalarità (se non un *continuum*) di possibili 'forme' della derivazione per mezzo di una scelta discreta tra un numero limitato di opzioni (che abbiamo identificato come il criterio della trasparenza semantica e il criterio dell'autonomia morfo-sintattica della base).

L'utilità del prodotto finale del progetto può però essere valutata anche su altri fronti. Sul versante, infatti, prettamente sperimentale, il CoLFIS, che - come abbiamo visto - già oggi in quanto dizionario di frequenza lessicale soddisfa i due fondamentali requisiti di rappresentatività del campione e bilanciamento delle fonti (requisiti che lo rendono strumento irrinunciabile per gli studi psicolinguistici sull'italiano), una volta corredato di informazioni morfologiche e prosodico-sillabiche potrà essere considerato il punto di riferimento oggi mancante per qualsiasi studio sperimentale che abbia

bisogno di informazioni statistiche affidabili sui maggiori parametri di differenziazione formale del lessico italiano.

Riferimenti Bibliografici

Bertinetto, Pier Marco & Burani, Cristina & Laudanna, Alessandro & Marconi, Lucia & Ratti, Daniela & Rolando, Claudia & Thornton, Anna Maria (2005). *Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS).*

Bortolini, U. & Tagliavini, C. & Zampolli, A. (1971). *Lessico di frequenza della lingua italiana contemporanea*, Milano: Garzanti.

http://linguistica.sns.it/CoLFIS/CoLFIS_home.htm

Cosi, Piero & Roberto, Gretter & Fabio, Tesser (2002). Festival parla italiano, in *Atti delle XI Giornate del Gruppo di Fonetica Sperimentale (GFS)*, Padova 29-30 Novembre - 1 Dicembre, 2000.

Gaeta L. (2002). *Quando i verbi compaiono come nomi: un saggio di Morfologia Naturale*, Milano: Franco Angeli.

Grossmann, Maria & Rainer, Franz (cur.) (2004). *La formazione delle parole in italiano*, Tübingen, Niemeyer.

Laudanna, Alessandro & Thornton, Anna Maria & Brown, G. & Burani, Cristina & Marconi, Lucia (1995). Un corpus dell'italiano scritto contemporaneo dalla parte del ricevente, in S. Bolasco, L. Lebart e A. Salem (cur.), *III Giornate internazionali di Analisi Statistica dei Dati Testuali*, Volume I, Roma, Cisu, pp.103-109.

Rainer, F. (2004). Tipologia delle restrizioni, in M. Grossmann & F. Rainer (cur.), 20-22.

Scalise, Sergio (1995). La formazione delle parole, in Renzi, L., Salvi, G. e Cardinaletti, A. (a cura di), *Grande grammatica italiana di consultazione*, vol III, Bologna, Il Mulino: 473-516.

Scalise, Sergio (1994). *Le strutture del linguaggio. Morfologia*, Bologna, Il Mulino.