

Approssimazioni al VoDIM

Manuel Barbera

*Non dire otto se non l'hai nel cruscotto
Uovo di sera frittata si spera*

Pseudo-proverbi da Benito Jacovitti, *Cocobill contro chissà?*,
in «Corriere dei piccoli» 24 (14 giugno 1970) - 27 (5 luglio 1970),
ora in *Cocobill e il meglio di Jacovitti*, a cura di Luca Boschi, Milano,
Hachette Fascicoli, 2018, vol. 35, pp. risp. 14 e 23.

Introduzione

Il VoDIM (Vocabolario Dinamico dell'Italiano Moderno¹), è un progetto, capitanato dall'Accademia della Crusca, che coinvolge otto gruppi di ricerca di altrettante università italiane, fra cui anche il gruppo torinese, volto a sviluppare un dizionario dell'italiano postunitario online, basato su corpora e su altri dizionari acquisiti in formato digitale come il Tommaseo-Bellini, la quinta Crusca ed il Battaglia, e disegnato per poter essere interrogabile anche a “corpus variabile”, definito dall'utente.

I corpora su cui si appoggia diventano quindi essenziali.

Un primo corpus di riferimento base (i cui risultati non sono stati ancora resi pubblici: <http://dizionariodinamico.it/prin2012crusca/dictionary>) è stato prodotto col PRIN 2012 dalla medesima Crusca (in collaborazione con le Università di Catania, Firenze, Genova, Milano, Napoli, Piemonte Orientale, Tuscia e con il CNR), ma, naturalmente, da solo è insufficiente alla bisogna. Altri corpora lo stanno affiancando, ed altri progetti lessicografici tra cui il VoSCIP (Vocabolario Storico della Cucina Italiana Postunitaria)².

Il presente contributo, prendendo le mosse, e pure terminando, col problema dei corpora, intende suggerire alcune strategie per l'organizzazione delle voci.

1. I NUNC

Oltre a questi corpora, un altro che pensavamo³ potrebbe utilmente completarli è il NUNC-IT. I NUNC (Newsgroups UseNet Corpora⁴), ideati da Manuel Barbera (in bmanuel.org), ed appannaggio del medesimo gruppo torinese che partecipa al VoDIM, propriamente sono una suite multilingue

1 Cfr. online: <http://www.accademiadellacrusca.it/it/eventi/crusca-torna-vocabolario-lessicografia-dinamica-dellitaliano-post-unitario>.

2 Per una breve presentazione cfr. Malgarini *et alii* 2018.

3 Barbera - Marellò 2018.

4 Homepage: <http://www.bmanuel.org/projects/ng-HOME.html>.

Cita come:

Manuel Barbera, *Approssimazioni al VoDIM*, “Italiano digitale”, VII, 2018/4, pp. 138-158.

di corpora che vorrebbe documentare il genere testuale “newsgroup” all’inizio del terzo millennio; molte versioni ne sono state implementate (anche per tematiche specifiche), tutte reperibili dalla homepage; il risultato non è ancora del tutto soddisfacente⁵; pure, qualche uso può già esserne fatto⁶. Un newsgroup è un forum telematico a libero accesso, gratuito, disponibile su internet, che si manifesta nella forma di testi scritti, i post, inviati ad una “bacheca elettronica” mantenuta presso una rete di server (i newserver che costituiscono UseNet). Gli utenti del gruppo possono scaricare, leggere e rispondere ai post, costruendo catene (thread) di botte e risposte. I newsgroup sono articolati in una tassonomia precisa, ossia in un sistema di cornici argomentative che si chiamano “gerarchie”, a base geografico-nazionale e/o tematica. I vantaggi di questa base testuale per la linguistica dei corpora sono numerosi e sono stati trattati in Barbera 2007 e Barbera-Marello 2009; qui ci interessa in primo luogo il fatto che presentano una *Umgangssprache* assolutamente contemporanea, reale e molto variata per registri e temi.

Per quanto riguarda il VoDIM, molte voci, neologismi, tecnicismi, prestiti, ecc., non sono attestate nel corpus base della Crusca e quindi i NUNC potrebbero risultare un utile serbatoio di contesti.

Un ulteriore fattore che rende i NUNC apprezzabili per il linguista e il lessicografo attento all’uso è la dialogicità, che si intravede soprattutto negli esempi. È un fenomeno pervasivo nei NUNC, di solito declinato nei newsgroup come *quoting* (cfr. Barbera 2011 e Marello 2007). Ciò crea computazionalmente, è vero, alcuni problemi (ancora non del tutto risolti), dato che il fenomeno del testo ripetuto, se incontrollato, va inevitabilmente ad intaccare l’aspetto statistico, vanificando un semplice uso quantitativo dei corpora; però testualmente è un fenomeno di grande importanza, specie se valorizzabile, come nei NUNC, con la possibilità di potere allargare i contesti fino a 2000 parole. Comunque, la capacità dei newsgroup di fissare nello scritto usi eminentemente orali, di trasferire la fluidità dell’oralità ad uno speciale tipo di scrittura, costituendo una sorta di ponte tra i due media, può rivelarsi particolarmente importante per il VoDIM, proprio perché questi corpora registrano tendenze emergenti nella lingua italiana (sulla peculiarità diamesica di questo particolare tipo di “scritto-parlato” abbiamo sostato in Barbera - Marello 2009).

2. Un primo case study: Umwelt

Si veda ad esempio un prestito di ambito tecnico, il termine *Umwelt*.

Introdotta (in tedesco) dal biologo (estone, ma di famiglia tedesca del Baltico) Jakob Johann baron von Uexküll già nel titolo della sua importante opera del 1909 (*Umwelt und Innenwelt der Tiere*, che però è più spesso letto nella seconda edizione del 1921), è entrata presto nella tradizione filosofica (a partire da una recensione di Max Scheler del 1914⁷): usata da Heidegger in un suo corso del 1929-30, è diventata poi moneta corrente (tra gli altri) in francese con Gilles Deleuze, Maurice Merleau-Ponty e Jacques Lacan, nonché in italiano con Giorgio Agamben. Ma è usata soprattutto in testi di biologia,

5 I NUNC, non sono ancora perfetti; non a caso dieci anni fa il sempre compianto Adam Kilgarriff scriveva: «cleaning is a low-level, unglamorous task, yet crucial: [...]. To date, cleaning has been done in isolation (and it has not been seen as interesting enough to publish on). Resources have not been pooled, and it has been done cursorily, if at all» (Kilgarriff 2007, p. 149); ed in un decennio non è che la situazione sia cambiata molto (se non che purtroppo Adam non c’è più; i criteri che avevamo usato sono riassunti in Barbera - Colombo 2010). Comunque, i prototipi che sono stati messi online sono solo delle beta, ma la volontà di perfezionarli c’è: e non è da escludere che il VoDIM rappresenti l’occasione giusta per tentarlo.

6 Come dimostrato da alcuni degli interventi presenti in Barbera *et alii* 2007; in Costantino *et alii* 2009, per non citare che i primi utilizzi ce ne sono stati fatti dieci anni fa.

7 La recensione è spesso citata, ma non sono mai riuscito a vederla.

naturalmente, e poi in semiotica⁸, in cui è stato diffuso negli anni Sessanta da Thomas Albert Sebeok (né Sebeők Tamás)⁹, in quanto è alla base della moderna biosemiotica (cfr. Kull 2001).

Le difficoltà cui va incontro nella lessicografia italiana il dominio lessicale della biologia sono state ben rappresentate da Gualdo 2018, pp. 202-206; ed *Umwelt* non fa in questo eccezione; nel corpus di base del VoDIM inoltre il termine non è presente, così come in genere nei dizionari italiani, ma nei NUNC è ripetutamente attestato¹⁰. Riprendiamone almeno due campioni.

Per Gadamer comprendere l' 'esistenza - e qui c'è ancora Heidegger - significa prima di tutto pre-comprenderla , in quanto la comprendiamo con un linguaggio che non scegliamo , ma che , trascendentalmente , definisce già la realtà in cui ci muoviamo : l'Um-Welt , da un lato , e dall ' altro lato , il Mit-welt . Ma , Gadamer cerca di andare alla radice del movimento del pensiero del soggetto e tale origine sta nell ' esigenza di comprendere e farsi comprendere , cioè nel muoversi nell ' Umwelt e nel Mitwelt . Il fatto è che per Gadamer l' Altro è visibile solo con gli " occhi nostri " , cioè con ciò che " siamo " , con la nostra " identità " , il nuovo si dà solo nel familiare . E in un certo senso è così . L ' altro è ciò che mi disturba che mi inquieta perchè non riesco a ridurlo al mio mondo : è un'eccezione .

Quello precedente è un esempio tipico dell'uso tecnico-filosofico del termine, che non si discosta molto da quello che si potrebbe trovare nello spogliare manualmente i testi (e le traduzioni) di quella tradizione. Più interessante è però l'esempio seguente:

Anche in Italia il consumo di televisione è vertiginosamente aumentato : [...] . Oltre a due effetti di rilevanza individuale : - la caduta verticale della capacità di fissare l' attenzione per più di un certo tempo (se a un buon insegnante occorre anche un ' ora per sviluppare un dato argomento , gli spazi televisivi obbligati in novanta secondi troncano quello stesso argomento in modo irreparabile) e - la perdita di interesse per la lettura - aspetti che coinvolgono per mimetismo inconscio (vale a dire per l' inconscio occupazione degli spazi mentali ad opera non solo delle immagini ma dell ' intera atmosfera televisiva che foggia l ' Umwelt dell ' uomo moderno) anche persone che fruiscono della TV per tempi ben sotto la media - l ' esposizione allo " sbarramento " delle immagini" televisive ha due rilevanti effetti sociali : - il conformismo applicato e - l ' ignoranza generalizzata. [...]

Si tratta di un uso traslato, chiaramente fuori dai campi "tecnici" di diffusione del termine. Lessicograficamente ciò è particolarmente rilevante perché testimonia il traghettamento del prestito al di fuori del dominio originario di appartenenza, assicurandone lo sdoganamento all'uso comune, anche se colto o relativamente tale.

3. Le future voci del VoDIM

Malgarini, Biffi e Vignuzzi nell'articolo citato (Malgarini *et alii* 2018, pp. 94-95) danno, riprendendola dal loro lavoro al VoSCIP (Vocabolario storico della cucina italiana postunitaria¹²), il seguente modello di schema per una voce-tipo:

8 Per la sua diffusione anche in altri settori cfr. Pollmann 2013.

9 La trama è stata dapprima sdipanata in una pretesa "recensione" apparsa su «Language» all'inizio degli anni Sessanta (Sebeok 1963; il termine *zoosemiotics* credo nasca a p. 465); per gli sviluppi della zoosemiotica, cfr. Martinelli 2010.

10 Le citazioni dal corpus nel prosieguo sono riportate immutate: in particolare si sono mantenute le tokenizzazioni di interpunzioni ed apostrofi, tutti gli "errori di digitazione", e le idiosincrasie ortografiche proprie del genere.

11 Come si diceva, qui le citazioni dal corpus sono state mantenute intatte, con tutti gli errori presenti nella fonte. Tantopiù che la maggiore tolleranza alle cattive digitazioni, e l'aperta accettazione di alcune caratteristiche grafico-ortografiche, sono tipiche di questo genere di CMR.

12 Cfr. <http://www.accademiadellacrusca.it/it/attivita/litaliano-cucina-per-vocabolario-storico-gastronomia>; cfr. anche Vignuzzi - Malgarini 2017.

LEMMA + categoria grammaticale

0.1. Forme attestate nel corpus dei testi (con tutte le varianti)

La forma lemmatizzata per la voce principale è quella più diffusa nell'uso odierno: ci si serve del GRADIT, *Grande dizionario italiano dell'uso*, di Tullio De Mauro, con i relativi aggiornamenti.

0.2. Nota etimologica essenziale.

0.3. Prima attestazione nel corpus.

0.3.1. Indicazione numerica della frequenza (per ciascuna forma; nell'indicazione delle occorrenze, la seconda cifra, preceduta dal segno +, si riferisce alle forme presenti in eventuali indici).

0.4. Distribuzione geografica delle varianti.

Per ora si forniscono i dati relativi ai soli AIS e ALI. Aggiungiamo in nota il riscontro con le forme registrate da Touring Club Italiano 1931.

0.5. Note linguistiche/merceologiche (forestierismi; italianismi in altre lingue).

La bibliografia per ora si riferisce solo alle 'Note linguistiche', e, per quanto riguarda gli italianismi in altre lingue, al DIFIT (consultabile in versione elettronica in <http://www.italianismi.org/difit-elettronico>).

0.6. Riepilogo dei significati.

0.7. Locuzioni polirematiche e proprie (con la prima attestazione nel corpus).

0.8. Rinvii (sono previsti soprattutto 'iperlemmi', o, se si preferisce voci 'generali', di raccordo).

0.9. Corrispondenze lessicografiche (= riscontri nei dizionari e nei corpora lessicografici in rete): si distinguono i vocabolari etimologici (compreso il LEI) da quelli descrittivi (in ordine cronologico, a partire dal Tommaso-Bellini).

1. Prima definizione

Contesti

1.1. Definizione subordinata

Contesti

1.2. Definizione subordinata

Contesti

[...]

2. Seconda definizione

Contesti

[...]

L'indubbio vantaggio di questa griglia è quello di essere sostanzialmente conforme a quella sottesa al TLIO (Tesoro della Lingua Italiana delle Origini¹³), prodotto dall'Istituto dell'Opera del Vocabolario Italiano del CNR di Firenze. Non mancano però anche gli svantaggi: gli autori del VoSCIP sottolineano soprattutto l'ipertrofia dello schema («la struttura rischia però di essere troppo pesante in vista di una effettiva fattibilità realizzativa del progetto», Malgarini *et alii* 2018 cit., p. 95), obiezione cui concorriamo. Ma v'è altro.

Una prima osservazione affatto generale è che questo suggerimento non offre alcuno spunto su come distinguere le diverse entrate, cioè per come affrontare il problema degli omografi: la decisione di quando si debbano avere voci distinte, piuttosto che sottovoci all'interno del medesimo lemma, è una di quelle questioni in cui la soggettività gioca, ed ha giocato, un ruolo assai rilevante. E ciò è chiaramente da evitare, anche se la faccenda non è di semplice soluzione: una discussione orientativa, fatta in anni ormai abbastanza lontani, si trova in Barbera 2009, pp. 707-712, dove la dimensione del fenomeno era al contempo più limitata (si trattava solo di costruire il lemmario per un corpus) e più grave (si trattava pur sempre di italiano antico); là suggerivamo la combinazione di due criteri: «(1) differenza semantica sostanziale e (2) almeno parziale differenza etimologica o di formazione morfologica» (Barbera 2009 cit., p. 707). Dei due il secondo è l'unico completamente oggettivo, ma da solo è ahimè il meno efficace: a fronte dei molti casi tipo *canto* 'canzone' vs 'angolo', di etimo ovviamente di-

13 Homepage: <http://tlio.ovi.cnr.it/>.

verso, non mancano casi come *prigione* ‘carcere’ vs ‘prigioniero’ con lo stesso etimo (PRĒHĒNSIŌNEM, REW 6737) ma significato distinto; infatti nell’ultimo caso la maggioranza dei dizionari (ad esempio il DISC, lo Zingarelli ed il Devoto - Oli) presenta due entrate lessicali distinte e praticamente solo uno ha voce unica (il Cortelazzo-Zolli, non a caso un etimologico; non però così il DEI). Purtroppo la “differenza semantica sostanziale” non è rigorosamente quantificabile, ma non è possibile sbarazzarsene del tutto; posso solo consigliare di limitarne i danni, invocandola solo quando il criterio etimologico fallisca il colpo o dia (che spesso è lo stesso) risultati dubbi.

Una seconda osservazione, collegata alla precedente, è che non sembra essere stata implementata una organizzazione delle voci a base POS (*Part of speech*) privilegiando una griglia di struttura a base semantica; ciò equivale (come peraltro quasi esplicito in questo schema) a scaricare queste distinzioni nell’entrata lemmatica, in pratica ponendo un terzo criterio (che, nell’ottica prima discussa, dovrebbe essere il primo ad agire) per generare voci omografe. Il che si può benissimo fare, anzi in molti casi potrebbe semplificare dei problemi di soluzione altrimenti opinabile.

Se la principale base organizzativa degli articoli è dunque semantica, bisognerebbe proporre qualche criterio organizzativo, per evitare, specie nelle voci più estese, la farragine, la gratuità e la difficile gestibilità, difetti che talvolta affliggono, ad esempio, il Battaglia.

Alcune categorie, infine, che erano senz’altro utili nel TLIO non lo sono forse più nel VoDIM: ad esempio “riepilogo” in o.6 serviva nell’italiano antico, dove spesso c’è una selva (quando non una foresta equatoriale) di variazione ortografica, ma è assai meno necessario nell’italiano moderno in cui i modelli ortografici sono più definiti: quali che siano le decisioni che verranno prese, qui se ne è fatto a meno.

4.0 Il case study di prima trasformato in voce: i forestierismi

Se torniamo al nostro *Umwelt* e proviamo a farne una voce, i problemi non mancano, anche a parte le considerazioni che abbiamo sviluppato nel paragrafo precedente.

Preliminarmente, quando si tratta di forestierismi¹⁴, bisogna dotarsi di un criterio per distinguere i prestiti non adattati (da prendere in un dizionario) dai termini stranieri *tout court* (da scartare in un dizionario). La nostra posizione era stata illustrata in Barbera 2015, p. 143; stante le regole seguenti,

- (1) la presenza *de facto* di una voce di origine straniera in un lessico specialistico;
- (2) il suo uso e frequenza anche fuori dal singolo dominio specialistico di partenza
 - (a) nella lingua parlata usuale,
 - (b) in più domini specialistici;
- (3) la presenza di derivati a morfologia italiana e la loro diffusione
 - (a) in condizioni del tipo (2),
 - (b) in condizioni del tipo (1);
- (4) la diffusione internazionale del prestito

applicate come illustrato¹⁵ nell’articolo citato, *Umwelt* deve essere accettato e pertanto avrà una voce. Naturalmente, una ovvia ricaduta di ciò, che comunque era anche altrimenti assai raccomandabile, è che bisogna essere molto attenti a distinguere ed indicare i diversi domini lessicali.

¹⁴ Sui prestiti in generale, e sui rapporti dell’italiano con l’inglese in particolare, la nostra prospettiva è la medesima delineata in Sgroi 2018. In questa ottica, il dominio lessicale della biologia è doppiamente importante perché l’asse di base della manualistica universitaria non è in italiano.

¹⁵ «La condizione (1) non basta da sola a far “accettare” un prestito (che pertanto resta termine straniero, in corsivo e con genere e pluralizzazione della lingua d’origine), anche se si potrebbe restare più possibilisti in casi in cui fosse (α) di alta frequenza, (β) inevitabile a pena di perdita di esattezza tecnica o concisione espositiva, e (γ) proprio del dominio specialistico che uno sta monograficamente utilizzando. La soddisfazione della condizione (2) da sola rende il prestito automaticamente accettato nel caso (a) e perlomeno possibile se non probabile nel caso (b); nella cui eventualità i fattori (α) e (β) *supra* possono essere dirimenti. Anche la soddisfazione contemporanea delle condizioni (2) e (3), cioè la situazione che per

4.1 Minimalmente

Quella di *Umwelt*, tanto per iniziare, sarà una voce “breve” per cui pochi dei campi predisposti possono essere usati. In prima istanza avremo dunque qualcosa del genere:

Umwelt s.m.¹⁶

0.1. Umwelt - Um-Welt

0.2. Prestito dal tedesco, da Uexküll 1909

0.3. Inizio 2000, NUNC

0.3.1. Umwelt 3 - Um-welt 1

1. Il fondamento biologico soggettivo di ogni specie, diverso dall'ambiente oggettivo condiviso da tutte le specie (biologia)

1.1. Il mondo in cui si dà la nostra esistenza (filosofia)

Per Gadamer comprendere l'esistenza - e qui c'è ancora Heidegger - significa prima di tutto pre-comprenderla, in quanto la comprendiamo con un linguaggio che non scegliamo, ma che, trascendentalmente, definisce già la realtà in cui ci muoviamo: l'Um-Welt, da un lato, e dall'altro lato, il Mit-welt. Ma, Gadamer cerca di andare alla radice del movimento del pensiero del soggetto e tale origine sta nell'esigenza di comprendere e farsi comprendere, cioè nel muoversi nell'Umwelt e nel Mitwelt. (NUNC 1)¹⁷

1.2. Ambiente

Anche in Italia il consumo di televisione è vertiginosamente aumentato: [...]. Oltre a due effetti di rilevanza individuale: - la caduta verticale della capacità di fissare l'attenzione per più di un certo tempo [...] e - la perdita di interesse per la lettura - aspetti che coinvolgono per mimetismo inconscio (vale a dire per l'inconscio occupazione degli spazi mentali ad opera non solo delle immagini ma dell'intera atmosfera televisiva che foggia l'Umwelt dell'uomo moderno) anche persone che fruiscono della TV per tempi ben sotto la media - l'esposizione allo “sbarramento” delle immagini televisive ha due rilevanti effetti sociali: - il conformismo applicato e - l'ignoranza generalizzata. (NUNC 1)

In primo luogo, le raccomandazioni di Malgarini, Biffi e Vignuzzi (Malgarini *et alii* 2018) dichiaravano che «la forma lemmatizzata per la voce principale è quella più diffusa nell'uso odierno: ci si serve del GRADIT, *Grande dizionario italiano dell'uso*, di Tullio De Mauro, con i relativi aggiornamenti»; ma ciò non è molto utile per voci generalmente assenti nella lessicografia precedente, come appunto questa. Restano allora due sole opzioni: o si ricorre alla veste che ha nell'originale, ma se si tratta di un forestierismo, la tentazione di ricorrere alle traduzioni può innescare un cortocircuito teoricamente infinito tra pratiche traduttive e pratiche lessicografiche; o si ricorre ad una fonte esterna, che sia sempre disponibile ed il più ricca possibile. Se, come noi crediamo, la seconda ipotesi è la più verosimile, addivenire alla Wikipedia¹⁸ è pressoché inevitabile: peraltro, stante la prevista pubblicazione online del VoDIM, automatizzare in HTML la presenza di link ipertestuali non è difficile. In tal caso sarebbe opportuno stabilire preliminarmente una gerarchia di opere di riferimento cui ricorrere. V'è anche un secondo problema: se i due usi traslati (discorso filosofico e discorso comune di stile alto: 1.1 e 1.2) sono attestati nei NUNC, non lo è l'uso proprio (1.), biologico, da cui pure questi deb-

brevità chiamiamo (3a), rende l'accettazione del prestito automatica, mentre la simultanea soddisfazione di (1) e (3), cioè (3b), rende l'accettazione del prestito perlomeno possibile, e la sua accettabilità è tanto più alta quanto più alta è la frequenza della base e dei suoi derivati e la potenziale diffusione dei derivati fuori del dominio specialistico di partenza. Ciò perché considererei già la creazione di una “famiglia lessicale” prova in sé dell'acclimatazione della base allotria nel lessico ospite. Nei casi più incerti, infine, sarà il fattore (4) a far pendere l'ago della bilancia da una parte o dall'altra» (Barbera 2015 cit., p. 143).

16 Uso *s.m.* per *default* essendo il sistema di abbreviazioni più frequente nella lessicografia italiana; confesso però la mia sostanziale indifferenza alla scelta di un *label* piuttosto che un altro, purché il *tag* cui si riferisce sia stabile e definito: l'uso di un nome per un *taxon* non è tanto rilevante quanto la definizione del sistema di *taxa* cui appartiene. Per i sistemi di POS (e la definizione dei *tags*) cfr. Barbera 2001.

17 Qui e nel prosieguo normalizziamo almeno la tokenizzazione.

18 La distinzione tra dizionario ed enciclopedia è d'altronde una delle più difficili da tracciare teoricamente: la letteratura in proposito si spreca, e non mancano esempi di pratiche dichiaratamente “miste” come quella dei Larousse.

bono partire. Il ricorso sistematico alla Wikipedia permetterebbe di trattare anche questo problema. Anzi, meglio ancora sarebbe se una definizione (o meglio, una “iperdefinizione” soprastante le definizioni secondarie), venisse data in testa, eventualmente appoggiata alla fonte primaria.

Un terzo problema è che non è previsto alcun campo per la morfologia. Questo si spiega facilmente con il fatto che lo “zero” era articolato per i problemi di una lingua morta, ma le esigenze di una lingua viva sono diverse, come il problema della formazione del plurale, particolarmente impegnativo per i prestiti non adattati, mette chiaramente in evidenza. Di qui a proporre un campo “o.io” per questo genere di informazioni, il passo è facile; meno come saturarlo. In una prospettiva puramente corpus-driven, è ben possibile che le forme che servono non si trovino nei corpora, in una prospettiva “impuramente” corpus-based è difficile evitare le insidie del purismo¹⁹.

Strutturalmente, inoltre, questo tipo di informazione dovrebbe più verosimilmente precedere tutte quelle poste in “zero” che seguirle: meglio quindi postulare uno “o.o” che uno “o.io”.

Inoltre, in quest’ottica di rimaneggiamento, il “3.I” dovrebbe più utilmente diventare un “1.I” riavvicinando così le cifre alle forme che computano.

4.2 Meno minimalmente.

Secondo quanto abbiamo delineato, in questa ipotesi la voce potrebbe prendere la veste seguente²⁰:

Umwelt s.m. «Secondo Jakob von Uexküll, Max Scheler e Thomas A. Sebeok, *umwelt* (plurale: *umwelten*; la parola tedesca *Umwelt* significa ‘ambiente’ o ‘mondo circostante’) è ‘il fondamento biologico che sta nell’esatto epicentro della comunicazione e del significato dell’animale-uomo (e non)’. Il termine è in genere tradotto con ‘universo soggettivo’. Uexküll ha teorizzato che gli organismi hanno diversi *umwelten*, anche se condividono lo stesso ambiente» (Wikipedia, s.v.)

o.o Plurale ?²¹

o.I. *Umwelt - Um-Welt*

o.I.I. 3 - 1

o.2. Prestito dal tedesco, da Uexküll 1909

o.3. Inizio 2000, NUNC

o.9. Wikipedia (con collegamento ipertestuale²²)

1. Il fondamento biologico soggettivo di ogni specie, diverso dall’ambiente oggettivo condiviso da tutte le specie (biologia)

19 La presenza di derivazione con formanti italiani è indice dell’avvenuto adattamento di un prestito, l’invariabilità nel numero va invece modulata: sacrosanta per l’inglese e l’ebraico (già tale in latino: *quis mihi tribuat ut ego moriar pro te Absalom fili mi fili mi*, si lamentava Davide secondo il salmista girolamino, II Sm 18, 33; Gryson 2007, p. 444b) come per l’arabo, il finlandese e l’ungherese, non lo è per il greco (le *poiesis*), il latino (i *corpora*) e forse il tedesco (i *lieder*). Spesso entrambe le forme (flessa come nella lingua d’origine, ed invariabile) sono attestate, anche se un uso tendenziale (come tale accettato in Barbera 2015, p. 144) sembra potersi ravvisare: preferire gli invariati per le parole da lingue moderne ed i declinati per termini dalle lingue classiche e dal tedesco.

Va peraltro ammesso che se per la distinzione tra prestito e parola straniera *tout court* (cosa che per stabilire il lemmario di un dizionario è primaria) i criteri sopra accennati permettono (almeno credo) di tracciare un discrimine abbastanza sicuro, altrettanta certezza non v’è sul confine tra prestito adattato e non adattato. I requisiti di fondo sono essenzialmente due: [1] la derivazione (di solito considerato *shibboleth* infallibile) e [2] la pluralizzazione; ma non solo non danno sempre risultati, ma a volte entrano anche in contraddizione tra di loro: l’anglismo *bar* dal punto di vista della pluralizzazione è invariabile (per *two bars* si ha normalmente *due bar* e non **due bari* – lasciamo perdere i vernacolari *barri* del toscano, che ha pure un singolare *barre* – o **due bars*), ma si possono trarne derivati come *baruccio* o *baretto*: quindi in base a [1] sarebbe adattato ed in base a [2] non adattato. Ed anche all’interno dello stesso criterio non c’è sempre uniformità: considerando questa volta i russismi, *vodka* (che pure mantiene la <k>) ha un plurale *vodche* (es. *cameriere, ci porti due vodche*), ma *samizdat* rimane comunque invariabile (e non si ha **ty* o **ti* per *-mь*). Abbisognerebbe invero uno studio apposito, in prospettiva descrittiva (meglio se corpus-based) e non puristica, del problema.

20 Con gli esempi in versione “normalizzata”, di cui già avevamo fatto un assaggio nella versione precedente della voce.

21 Nei corpora non è attestato; in base al criterio sopra enunciato si avrebbe *umwelten*, ma preferisco rifuggire dall’ignominia di essere tacciato di purista.

22 URL: <https://it.wikipedia.org/wiki/Umwelt>.

1.1. 'Il mondo in cui si dà la nostra esistenza' (filosofia)²³.

Per Gadamer comprendere l'esistenza - e qui c'è ancora Heidegger - significa prima di tutto pre-comprenderla, in quanto la comprendiamo con un linguaggio che non scegliamo, ma che, trascendentalmente, definisce già la realtà in cui ci muoviamo: l'Um-Welt, da un lato, e dall'altro lato, il Mit-welt. Ma, Gadamer cerca di andare alla radice del movimento del pensiero del soggetto e tale origine sta nell'esigenza di comprendere e farsi comprendere, cioè nel muoversi nell'Umwelt e nel Mitwelt. (NUNC 1)

[Es. 1]

1.2. 'Ambiente'

Anche in Italia il consumo di televisione è vertiginosamente aumentato: [...]. Oltre a due effetti di rilevanza individuale: - la caduta verticale della capacità di fissare l'attenzione per più di un certo tempo [...] e - la perdita di interesse per la lettura - aspetti che coinvolgono per mimetismo inconscio (vale a dire per l'inconscio occupazione degli spazi mentali ad opera non solo delle immagini ma dell'intera atmosfera televisiva che foggia l'Umwelt dell'uomo moderno) anche persone che fruiscono della TV per tempi ben sotto la media - l'esposizione allo "sbarramento" delle immagini televisive ha due rilevanti effetti sociali: - il conformismo applicato e - l'ignoranza generalizzata. (NUNC 1)

[Ess. 3]

Il principale difetto di una struttura simile, a fronte di un maggiore perspicuità, è che è più dispendiosa di spazio; ma questa è una considerazione che in una pubblicazione online è di secondaria importanza.

Un ulteriore miglioramento sarebbe che nella visualizzazione finale la numerazione "articolata" scompaia lasciando solo quella delle sottovoci (ricavabile automaticamente dalla precedente a partire da "r"). La sua assenza anche sullo sfondo è tuttavia sconsigliabile, in quanto può essere di aiuto ai redattori per mantenere una salda omogeneità di struttura, e garantisce la rimappabilità con il TLIO, scopo che mi sembra tra le ragioni di fondo della scelta del VoSCIP. In tale ulteriore ipotesi la voce potrebbe assumere un aspetto del genere:

Umwelt s.m. «Secondo Jakob von Uexküll, Max Scheler e Thomas A. Sebeok, *umwelt* (plurale: *umwelten*; la parola tedesca *Umwelt* significa 'ambiente' o 'mondo circostante') è 'il fondamento biologico che sta nell'esatto epicentro della comunicazione e del significato dell'animale-uomo (e non)'. Il termine è in genere tradotto con 'universo soggettivo'. Uexküll ha teorizzato che gli organismi hanno diversi *umwelten*, anche se condividono lo stesso ambiente» (Wikipedia, s.v.)

Plurale ?

Umwelt - Um-Welt.

3 - 1

Prestito dal tedesco, da Uexküll 1909

Inizio 2000, NUNC

Wikipedia

1. Il fondamento biologico soggettivo di ogni specie, diverso dall'ambiente oggettivo condiviso da tutte le specie (biologia)

'Il mondo in cui si dà la nostra esistenza' (filosofia)

Per Gadamer comprendere l'esistenza - e qui c'è ancora Heidegger - significa prima di tutto pre-comprenderla, in quanto la comprendiamo con un linguaggio che non scegliamo, ma che, trascendentalmente, definisce già la realtà in cui ci muoviamo: l'Um-Welt, da un lato, e dall'altro lato, il Mit-welt. Ma, Gadamer cerca di andare alla radice del movimento del pensiero del soggetto e tale origine sta nell'esigenza di comprendere e farsi comprendere, cioè nel muoversi nell'Umwelt e nel Mitwelt. (NUNC 1)

[Es. 1]

'Ambiente'

Anche in Italia il consumo di televisione è vertiginosamente aumentato: [...]. Oltre a due effetti di rilevanza individuale: - la caduta verticale della capacità di fissare l'attenzione per più di un certo tempo [...] e - la perdita di interesse per

²³ Il sistema dei domini lessicali è un punto che andrebbe particolarmente curato, dato che è un settore in cui è facile peccare di estemporaneità, che difatti è inconveniente in cui spesso cadono anche i migliori.

la lettura - aspetti che coinvolgono per mimetismo inconscio (vale a dire per l'inconscio occupazione degli spazi mentali ad opera non solo delle immagini ma dell'intera atmosfera televisiva che foggia l'Umwelt dell'uomo moderno) anche persone che fruiscono della TV per tempi ben sotto la media - l'esposizione allo "sbarramento" delle immagini televisive ha due rilevanti effetti sociali: - il conformismo applicato e - l'ignoranza generalizzata. (NUNC 1)
[Ess. 3]

5.0 Problemi supplementari: un altro case study, alienato

Quello di *Umwelt* era un caso molto speciale: assai breve (e che pertanto neutralizza molte parti di quello schema di base), tecnico, ed assente nel corpus di base (obbligando dunque al ricorso, di rincalzo, ai NUNC), ci ha permesso però di meglio definire la zona, indubbiamente problematica, dei prestiti non adattati e di iniziare ad orientarci su qualche problema generale.

Se prendessimo un caso più "normale", cioè desumibile dal corpus di riferimento ed abbastanza ampio da essere significativo, ma non troppo ampio da diventare di difficile controllo, altri nodi verrebbero verosimilmente al pettine.

5.1 Alienato nel corpus del VoDIM

Cercando *alienato* anche come forma flessa²⁴ nel corpus del VoDIM troviamo 77 forme utili; ai risultati ottenuti vanno infatti, preliminarmente, sottratte alcune forme che richiedono aggiustamenti o che appartengono altrove. Scremati i 15 duplicati ed un esempio sicuramente erroneo²⁵, restano da eliminare un esempio ritagliato con contesto insufficiente per prendere qualsiasi decisione²⁶, ed una transcategorizzazione esterna (*alienate* 2a persona plurale del presente indicativo di *alienare*).

Anche così ristrettici alle sole 76 forme²⁷ in discussione, i problemi da affrontare certo non mancano. Tanto per iniziare molte forme (37: 27 *alienato* e 9 *alienati*) sono chiaramente participi verbali,

Se il donatario ha alienato i beni, deve restituirne il valore, avuto riguardo al tempo della domanda, e i frutti relativi, a partire dal giorno della domanda stessa (Regno d'Italia - Regio Decreto 16 marzo 1942, n. 262 - Approvazione del testo del Codice civile).

e vanno senza dubbio poste sotto il verbo *alienare* (come, però, vedremo, non è indifferente); appartengono prevalentemente (in 32 casi) all'accezione giuridica 'cedere, sottrarre' ed in soli 4 casi a quella generica di 'allontanare':

E mi sono anche alienato l'animo di Tommy (Giacosa, Giuseppe - Come le foglie).

In 8 casi abbiamo indubbiamente a che fare con aggettivi

²⁴ Cioè attivando l'opzione «Cerca anche come forma flessa».

²⁵ «Quando vi alienate, rispettate le regole del maneggio e non disubbidite all'allenatore. (Malaspina, Simonetta - Come si fa e come non si fa. Manuale moderno di galateo): lo sfuso è evidente e facilmente eliminabile.

²⁶ «venduto, alienato, ecc. Si sopprimano dunque le parole nessuna eccezzuata, le quali non hanno significazione, pel timore che coloro cui spetterà d'interpretare la legge, volendo dar loro qualche significazione, questa non sia contraria all'intendimento nostro. (Rattazzi - VIII Legislatura - Tornata del 13 gennaio 1862)». In realtà i casi di questo tipo sono forse tre, perché il problema investe anche «Contiguo alla stanza dell'alienata eravi un camerino destinato a guardaroba; (Caracciolo De' Principi di Fiorino, Enrichetta - Misteri del chiostro napoletano)», e «e le scoperte e le indagini di Jenner, di Corvisart, di Avenbrugger, di Loenner e di Pinnel, il redentore degli alienati, avevano rivoltata come un guanto la scienza. (Praga, Emilio - MEMORIE DEL PRESBITERIO SCENE DI PROVINCIA)» (peraltro pure duplicato), ma qui è facile risalire agli originali (che sono presenti in archive.org, rispettivamente: <https://archive.org/search.php?query=Enrichetta%20Caracciolo> e <https://archive.org/search.php?query=Emilio%20Praga>) e controllare i passi, scoprendo che si tratta sicuramente in entrambi i casi di sostantivi, come si supponeva. Che i contesti nel corpus vadano più curati è raccomandazione ovvia: attualmente sono peraltro molto variabili, passando dal ridottissimo al ridondante.

²⁷ Il sistema dichiara «89 occorrenze in 49 testi», che anche tenendo conto dei necessari correttivi (76+15+1+1), non è molto spiegabile; il meccanismo di computo forse andrebbe rivisto, perché così i suoi risultati rischiano di essere depistanti.

Così l'alienista alienato terminò con il peggiore morbo in una casa cosiddetta di salute (Faldella, Giovanni - Donna Folgore).

Ma in ben 11 casi (tutti appartenenti all'accezione giuridica 'ceduto'²⁸) la scelta tra aggettivo o participio è francamente indecidibile:

Miglioramenti e deterioramenti dell'immobile alienato (Regno d'Italia - Regio Decreto 16 marzo 1942, n. 262 - Approvazione del testo del Codice civile).

Il problema è sistemico (sicché si ripercuoterà in moltissime altre voci del dizionario), ben noto ed insolubile: a fianco di forme che sono sicuramente participi ("Giovanni ha *chiuso* la porta") e ad altre che sono aggettivi ("Giovanni ha trovato la porta *chiusa*") ve ne sono di quelle opache ("la porta è *chiusa*" che può voler dire tanto 'la porta è stata chiusa' quanto 'la porta ha la qualità di essere chiusa'). Una decisione in proposito va pertanto presa; e credo ci siano solo tre possibilità di base: (1) convenzionalmente le forme incerte sono sempre aggregate ad una POS scelta arbitrariamente; (2) si crea una terza POS V=A; (3) si scartano semplicemente le forme e non se ne tiene conto. Proporrei per ora di semplificarci le cose, visto che le difficoltà già non mancano, adottando la terza soluzione (possibilmente segnalando l'eventuale numero di casi eccedenti sulla voce autonoma dell'aggettivo): uno dei vantaggi dell'editoria online, d'altra parte, è la facilità di introdurre modifiche, sicché nulla ci vieta di ritornare in un futuro sui nostri passi.

Sgombrato così il campo restano 8 aggettivi e 36 nomi, che costituiscono i dati primari da organizzare. Le accezioni nei nomi sono compatte (v'è solo il senso di 'pazzo') mentre quelle degli aggettivi sono più sfrangiate (il senso più frequente – 4 esempi – è quello sociologico e filosofico di 'inautentico, reso dissociato dalla realtà', seguito da quello di 'pazzo' – 2 –, e poi da quelli giuridico di 'ceduto' e quello generico di 'deprivato').

5.2 Possibili voci

Se ora abbiamo definito la situazione, proviamo in prima battuta a guardarci intorno.

Il modello più importante, dal punto di vista del VoDIM, per una lingua in cui si accampi una situazione abbastanza confrontabile con quella italiana, è probabilmente quello del *Trésor* francese (TLF, ora in versione online TLFi). Nelle linee essenziali la sua voce corrispondente è così organizzata:

I. Part. Pass.

II. Adj.

A - chose => sens juridique 1804

B - personne

1. [En parlant d'une pers. considérée dans ses rapports avec elle-même] Devenu étranger à soi-même

a) PSYCH. Devenu fou 1913

b) PSYCHOL. [Sans idée de folie] Devenu autre, privé de sa personnalité 1922

²⁸ Potrebbe a questo proposito venire il sospetto che *cessione* ed *alienazione* non siano perfettamente identici come istituti giuridici; in effetti dalla lessicografia esistente per l'italiano si evince senza dubbio che nella lingua ordinaria lo sono, ma che lo siano anche nella lingua del diritto non emerge con altrettanta chiarezza; personalmente, per esserne sicuro, ho dovuto chiedere ad un avvocato... Questo pertiene più all'enciclopedia che al dizionario, si dirà: forse. Ma al di là di ciò, evidenzia una chiara pecca dello schema di voce suggerito: non prevede l'indicazione di sinonimi ed antonimi. Che invece, secondo me, andrebbero utilmente sempre messi, anche se limitati a non più di un paio (mica siamo il Tommaseo!), ma distribuiti per le varie accezioni. Col che casi ipoteticamente dubbi come quello precedente si dissolverebbero da soli. Resta però il problema di come acquisirli senza ricorrere al *iudicium* del lessicografo (od alla sua *competence*, che suona indubbiamente meglio): non è un caso che nelle voci proposte in questo contributo me ne sia da principio astenuto, per poi nelle versioni finali fare delle proposte di mia invenzione, a puro titolo esemplificativo.

- 2 [En parlant d'une pers. considérée dans ses rapports avec la nature ou avec autrui]
 a) Vieilli. Devenu moralement étranger, hostile 1798
 b) PHILOS., SOCIOL. [Surtout dans la philos. marxiste] Devenu esclave des choses ou d'autrui 1945

III. Emploi subst.

- A. PSYCH. Personne affectée d'aliénation mentale 1891
 B. PHILOS., rare. Personne devenue esclave 1957

Questa struttura è indubbiamente molto diversa da quella prima ipotizzata, già fin dalla griglia principale della voce, che è a base-POS e non a base semantica, sistema che è peraltro condiviso anche da molta lessicografia italiana; si veda ad esempio il DISC (sempre sintetizzato allo stesso modo),

agg 'ridotto in stato di alienazione'

s.m. 1 psych. 'malato di mente, pazzo' Sin. folle, squilibrato

2 sociologia 'Chi subisce un processo di alienazione.'

E part. pass. di alienare, sec. XIV

od il GRADIT (di cui riporto al completo la voce):

alienato /alje'nato| (a-lie-na-to) p.pass., agg., s.m.

1 p.pass. → alienare, alienarsi

2 agg., s.m. TS psic. → malato di mente | CO fig. che, chi si sente demotivato e frustrato perché svolge un'attività in modo ripetitivo e senza responsabilità

Derivati: ²abalienato, alienatamente

□ (19)

La struttura del GRADIT è francamente confusa. Ed il Battaglia, che mischia inestricabilmente semantica e POS, è ancor peggio:

Alienato (part. pass. di alienare), agg. Trasferito nel possesso di altri; venduto.

- Sm.

2. Alieno, allontanato, respinto.

3. Agg. e sm. Demente, pazzo, che non possiede più le facoltà mentali.

4. Astratto, estatico; fuori di sé, smarrito (nei sensi e nello spirito).

5. Secondo il marxismo, estraniato dai prodotti della sua attività e dall'attività stessa nella società capitalistica (una persona).

Ma a parte ciò, nel TLFi la griglia semantica, che interviene dopo quella per POS, è molto più gerarchica di quella "atomica" del modello che penseremmo per il VoDIM (e di cui più avanti proporrò qualche assaggio); inoltre, i due livelli sono piuttosto promiscui, e misto è pure l'uso della semantica abbinata al genere / dominio lessicale. A mio parere, una struttura complessa di questo tipo, per quanto ipoteticamente efficace, è meno facilmente gestibile in modo automatico, e quindi una griglia meno "impalcata" sarebbe preferibile, a patto che sia fatta con criterio; donde il modello "atomico" di cui dicevamo.

Ed alcune altre questioni a questo punto si affacciano. E tra le molte, almeno tre non possono non essere prese in considerazione.

Andrebbe in primo luogo deciso se una accezione tecnica (largamente maggioritaria²⁹) possa diventare il significato principale (nel DISC lo è per il verbo ma non per il N/Adj; e quanto agli altri sensi il DISC ha 1 solo per il V, mentre N/Adj hanno 2 e 3), e del pari se un'accezione traslata (e quindi *logicamente* secondaria) possa diventare primaria se la accezione-base fosse rara o non attestata. Nel caso dell'accezione tecnica sarei di parere favorevole, lasciando che fosse la frequenza nel corpus a gestire l'ordine³⁰: le altre opzioni rischierebbero di essere poco pratiche ed in alcuni casi (se spacciate come realistiche e non convenzionali) potrebbero mascherare dei falsi storici (è capitato che parole nascano come tecniche e solo poi diventino generiche). Le stesse considerazioni, nei confronti del traslato, mi farebbero propendere per l'ipotesi negativa, dato che le ragioni della logica vogliono bene la loro parte, e che questa è complessivamente più gestibile.

In secondo luogo ci si dovrebbe fornire di un criterio unico per decidere quale debba essere la voce principale cui le altre rimandano, non fosse che per l'etimologia. Anche qui la forbice è tra la verità storica (si pensi che anche solo la derivazione *de facto* non ha sempre seguito l'iter che la logica manualistica vorrebbe) e l'opportunità sincronica. Ci si può, penso, anche orientare su una miscela accorta (e soprattutto esplicitamente dichiarata) delle due componenti, il cui peso andrebbe deciso in base alle finalità globali del progetto.

Etimologia, si diceva; e questo ci conduce direttamente al terzo punto. Fermo restando che un dizionario puramente sincronico³¹ potrebbe anche non dare le etimologie piuttosto che darle male (cfr. Barbera 2016, p. 565), ha senso imitare la formula del TLFi e dare le prime attestazioni seguendo analiticamente ogni partizione e non presentarle in un unico blocco come nel DELI che ricalcava il TESz (che ne fu il grande modello, il *μεγαλόψυχος* padre della formula "storico-etimologica")? La risposta è duplice: (a), le prime attestazioni bisogna innanzitutto averle, e dato che il corpus del VoDIM è sostanzialmente recente, tranne che per i neologismi (tipo *Umwelt*), queste dovranno comunque essere ricavate da altri dizionari: quindi bisogna elaborare una strategia globale; (b) lo schema tipo TLIO che è stato proposto (cfr. § 3) prevede uno specifico campo per l'etimologia (o.2), e si potrebbe usare quello, tantopiù se è mantenuto distinto dalla derivazione morfologica.

5.3 Verso il VoDIM

Stante queste osservazioni, le voci possibili, divise in base alla POS, ed articolate internamente in base alla semantica (in modo frequenziale-atomico e non ramificato) potrebbero a grandi linee assumere questo aspetto, rappresentando anche la "criptonumerazione":

- Alienato*, s.m. 'malato di mente, pazzo'.
 o.o. Participio passato di *alienare*. Maschile. Plurale alienati³².
 o.1. *alienati* - *alienato*.
 o.1.1. 17 - 4.
 o.2. Da *alienare*.
 o.3. ?³³

29 Cioè più frequente negli esempi dei corpora di riferimento (qui quello del VoDIM), non prevalente nelle scelte dei lessicografi italiani e di altre lingue romanze.

30 Già questa considerazione da sola dovrebbe bastare a scongiurare l'impostazione proteiforme delle voci del Battaglia, la cui imprevedibilità strutturale è anche un serio ostacolo ad ogni ipotesi di acquisizione automatica di materiali, pur tentante data l'indiscutibile ricchezza della fonte.

31 Un *seletus-sōnaraamat* 'dizionario definitorio' come l'EKS, che infatti non reca né le etimologie né le prime attestazioni.

32 Questa volta il plurale è comunque attestato nel corpus.

33 Il campo o.3 lo ho lasciato vuoto, perché non mi pare che la «prima attestazione nel corpus» (viste le differenze tra il corpus del TLIO e quello del VoDIM) sia un dato particolarmente utile: lo sarebbe la prima attestazione assoluta (cfr. *supra*), ma di questa non si è ancora stabilita l'estensione (tema generale o singole forme derivate?) e la fonte (dove procurarsela, se il corpus, non avendo maggiore profondità dell'Ottocento, evidentemente non serve?).

o.8. alienato₂, alienare.

o.9. Corrispondenze lessicografiche sotto alienare?

1. 'pazzo'³⁴

In questo mondo di distratti, di egocentrici, di saccenti, di noiosi, di indifferenti, di alienati, un buon ascoltatore è un esemplare raro, ricercato e gradito, assai più del conversatore indefesso. (Gasperini, Brunella - Il galateo).

Disposizioni sui manicomi e sugli alienati. (Biancheri - X legislatura - Tornata del 7 aprile 1870).

[21 ess.]

*Alienato*₂ a. 'dissociato, ceduto, deprivato'

o.o. Partecipio passato di alienare.

o.1. alienati - alienato - alienata

o.1.1. 5 - 5 - 1 - 1

o.2. Da alienare.

o.3. ?

o.8. alienato₂, alienare³⁵.

o.9. Corrispondenze lessicografiche sotto alienare³⁶

1. 'inautentico, reso dissociato dalla realtà' (sociologico e filosofico)

Nel «Supermarket», invece, (che è opera più recente e che riflette un ordine di esperienze più «alienato») lo schema astratto chiude in una sorta di favola cattiva i carrelli della abbondanza, le ruote che una goccia di nero rende per sempre immobili, i triangoli e i rettangoli delle merci perfette, senza sapore e odore, miraggi e miti della epoca nostra. (Personaggi e vicende dell'arte moderna).

Secondo il rapporto Warren Jack Ruby era un uomo profondamente alienato dalla società in cui viveva e con alle spalle un passato di fallimenti e di miseria. («La Stampa» 2 (04-01-1967)).

[4 ess.]

2. 'pazzo'

Così l'alienista alienato terminò con il peggiore morbo in una casa cosiddetta di salute. (Faldella, Giovanni - Donna Folgore).

Questa tomba di vive fu fondata da una pinzochera, semi-alienata di mente, coll'approvazione e sotto il patrocinio della Chiesa romana. (Caracciolo De' Principi di Fiorino, Enrichetta - Misteri del chiostro napoletano)

[2 ess.]

3. 'ceduto' (giuridico) [+I?]

La novità dell'istituto concerne il diritto del creditore munito di titolo esecutivo di espropriare, senza dover esperire preventivamente l'azione revocatoria ordinaria (art. 2901 c.c.), il bene immobile o mobile registrato alienato a titolo gratuito dal debitore ad un terzo oppure fatto oggetto di un vincolo di destinazione. (AA.VV. - Anno 2016)

[1 es.]

4. 'deprivato' [+I?]

la forza poetica e morale di gran parte delle opere presentate» (è ben diversa dal) «generale tono urlante della pittura di oggi, così irrazionalmente tesa al successo, da sostituire al contenuto della vita moderna un triste contenuto alienato di ambizione di ufficialità (?) e di pseudo originalità a tutti i costi. (Personaggi e vicende dell'arte moderna).

[1 es.]

Le osservazioni da fare sono moltissime; manteniamoci all'essenziale.

Una prima è che la macrostruttura della voce può apparire esageratamente isolante là dove quella del TLFi era forse troppo complessa ed embriicata (il modello-TLIO prevedeva d'altra parte la possibilità di un doppio livello). Evitare il soggettivismo eccessivo nell'organizzazione delle voci avevamo detto che è una delle esigenze più rilevanti, ed il ricorso ad una seriazione gerarchicamente la più semplice possibile potrebbe esserne un modo (da testare, naturalmente, in voci più lunghe). Appoggiato dalla considerazione "dinamica" che, ponendo in seriazione per frequenza i significati resi "atomici", l'u-

34 I traducanti usati sono ovviamente provvisori, e nella loro sinteticità non hanno altra funzione che orientare.

35 Ho qui limitato al minimo, cioè a quelle usate nell'esempio, le altre voci collegate, presenti solo per salvare le funzioni esplicative.

36 Per ipotesi.

tente potrebbe variare oltre che il numero degli esempi, anche il numero dei traducanti, “tagliandoli” per numero di occorrenze.

Una seconda è che l'appoggio al GRADIT³⁷, programmaticamente dichiarato da Malgarini, Biffi e Vignuzzi per le “arcidefinizioni” e dintorni, qui non ha potuto funzionare, per la eccessiva distanza delle due strutture (che nel GRADIT sono miste come nel TLF; il che determina forti differenze anche nei lemmari); e visto che la divergenza è sistemica, ci si può aspettare che questo problema si verificherà assai spesso nella redazione del dizionario. Sicché un'altra soluzione è auspicabile (qui abbiamo agito da occasionali *bricoleurs*, che non è altrimenti molto proponibile).

Una terza questione è quella dei participi. Come preannunciato, si è provato ad annotare sugli aggettivi il numero delle forme indecidibili; ma le decidibili e sicuramente assegnate al participio? Dobbiamo per forza rassegnarci a perderle diluendole nelle acque del vasto mare verbale di *alienare*? Forse non del tutto, se si segue il suggerimento del TLFi e del GRADIT: anche senza fare una vera voce autonoma distinta da quella completa di *alienare*, si potrebbe fare una scheda di rimando con poco più dei meri dati quantitativi.

Nell'ottica di numerare le voci per frequenza, potremmo avere allora:

*Alienato*₁ part. ‘cedere, allontanare’.
 o.o Participio passato di *alienare*
 o.i. *alienato* - *alienati*.
 o.i.i. 32 - 4.
 o.2. Da *alienare*.
 o.9. vedi *alienare*.
 1. ‘ceduto, tolto’ (giuridico) [32 ess.]
 2. ‘allontanato’ [4 ess.]
*Alienato*₂ s.m. ‘malato di mente, pazzo’.
 [...]

*Alienato*₃ a. ‘dissociato, ceduto, deprivato’
 [...]

La “dinamicità” potrebbe anche suggerire una ricerca “telescopica”, per cui si potrebbe dare in prima istanza una schermata del solo scheletro, poi ampliabile nelle direzioni volute, tipo

*Alienare*³⁸ v.t.³⁹ ‘allontanare, cedere’.
*Alienarsi*₁ v.t.p.⁴⁰ ‘allontanare da sé’.
*Alienarsi*₂ v.i.p. ‘distogliersi’.
*Alienato*₁ part. ‘cedere, allontanare’.
 Participio passato di *alienare*

37 E quanto al GDLI, i cui materiali essendo già stati acquisiti in Crusca ha certo senso usare, non va affatto meglio: si veda cosa avevamo già osservato in precedenza.

38 Le supposte voci verbali sono, appunto, solo supposte e poste qui a mero titolo esemplificativo e senza impegno alcuno: nessuno spoglio specifico è stato condotto. Il quadro derivazionale è volutamente stato dato in modo assolutamente incompleto.

39 La scelta di un sistema abbreviativo piuttosto che un altro è del tutto convenzionale: come già avevo detto sono abbastanza indifferente alla scelta del *label*, non a quella del tagset.

40 La decisione di avere voci distinte per pronominali, riflessivi ecc., oltre che per transitivi ed intransitivi (ed i meno tradizionali inaccusativi dove li mettiamo?) è un'altra di quelle questioni che richiederebbero una disamina approfondita (e forse monografica). Per semplificare, in sede generale (cfr. ad es. Barbera 2009, pp. 631-632) e teorica, sono sempre stato contrario all'idea di considerarle categorie autonome e di avere voci separate, ma qui potrebbero subentrare istanze diverse: tanto l'esigenza (a livello di macrostruttura) di razionalizzare l'articolazione del lemmario, appoggiandosi più che alla semantica ogni qual volta possibile alla divisione per categoria grammaticale, sempre “oggettiva”, tanto quella (a livello di microstruttura) di rendere agile ed “atomica” l'articolazione interna delle voci, questa volta invece base semantica, entrambe concorrerebbero a favorire la soluzione distinta. A mero titolo di discussione, qui ho provato ad implementare la struttura a voci separate.

1. 'ceduto, tolto' (giuridico)
 2. 'allontanato'
- Alienato*₂ s.m. 'malato di mente, pazzo'.
*Alienato*₃ a. 'dissociato, ceduto, deprivato'
1. 'inautentico, reso dissociato dalla realtà' (sociologico e filosofico)
 2. 'pazzo'
 3. 'ceduto' (giuridico)
 4. 'deprivato'

Da questa disamina restano scoperte le collocazioni (e gli *idioms* ad esse collegati), perché meriterebbero una trattazione a parte, essendo un capitolo troppo vasto per essere affrontato in questa sede. A grandissime linee, qui basta dire che il loro trattamento potrebbe essere duplice: in parte alcune vanno scaricate sul lemmario ed in parte altre vanno poste all'interno delle voci, sotto ogni partizione primaria, che così si arricchirà oltre che dei sinonimi e degli antonimi anche delle collocazioni. Il problema base sarà quello di fornirsi di regole rigide e non soggettivamente mutevoli per compiere le scelte appropriate.

6. *Micorriza e simili*

Abbiamo in questa disamina più volte incontrato i limiti (che in qualche misura sono inevitabili) del corpus di base del VoDIM, anzi avevamo proprio principiato da questi, ed ora a questi torniamo. Un'opera lessicografica che voglia essere completamente *corpus driven*, come l'impresa sinclairiana del COBUILD, ammesso che ciò sia davvero auspicabile, richiede perlomeno dei corpora di base ben bilanciati e pressoché perfetti o comunque adeguati. Condizione, purtroppo, da cui siamo, e verosimilmente resteremo, abbastanza lontani, sicché dobbiamo premunirci, predisponendo (e il più possibile automatizzando) il ricorso a delle fonti lessicografiche secondarie. Cosa che era peraltro già prevista nel progetto iniziale, ma che questi affondi confermano saldamente.

Un modo diretto, infatti, per toccare con mano questa realtà, è constatare l'assenza di molte parole che resterebbero così al di fuori. Si tratta spesso di sicure parole dell'uso moderno, non di quegli arcaismi che si trovano giustamente solo nel Battaglia, tipo *gramuffa* (GDLI VI, p.1931b) e *strabevizione* (GDLI XX, p. 246a), o neanche in quello (come il famoso *apostelesmatico* presente nella *Filosofia antica* dell'Adorno ma non altrove).

Un esempio potrebbe essere *micorriza*, per tornare al campo biologico (qui micologico e botanico) che mi è caro e che ho fatto vedere (*supra*) essere particolarmente "sensibile"; la sua presenza nel GRADIT costituisce già da sola un efficace passaporto ed una garanzia di sdoganamento nell'uso concreto (siamo cioè più verso la *benzodiazepina*⁴¹ che non verso il *desmetildiazepam*). Nel GRADIT si ha:

micorriza /mikor'ridza] (mi-cor-ri-za) s.f. TS Bot. ⁴²

[1957; comp. di *mico-* e *-rriza*]

complesso simbiotico costituito dal micelio di un fungo e dalle radici o rizomi di piante, le quali ricevono acqua e sali assorbiti dal terreno e cedono al micelio i carboidrati che hanno elaborato

Derivati: ectomicorriza, endomicorriza, micorizzare, micorrizico, micorizzazione, micorriza, micorizzogeno]⁴³

□ (19)

⁴¹ Presente infatti nel GRADIT, anche con il derivato *benzodiazepico*.

⁴² Si ricordi quello che si diceva a proposito dell'ottimizzazione e della standardizzazione del sistema dei domini lessicali: che la micologia faccia parte della botanica è cosa che dovrebbe fare inorridire anche più del considerare le cozze pesci.

⁴³ Integro i derivati presenti nel lemmario a quelli dichiarati dalla voce per far vedere il livello di integrazione del termine.

Cosa fare di queste parole (a volte, come questa, assenti anche nei NUNC), ammesso che le riuscissimo sempre a “scoprire” non essendovi nel corpus? Le strade sono essenzialmente due: o migliorare il corpus rendendolo *à tout faire*, come quello del COBUILD, ma è molto difficile⁴⁴, o standardizzare il ricorso ad una o più fonti esterne.

E temo che la seconda strada (anche se resa il più marginale possibile), come altrimenti programmato, sia davvero inevitabile. Si potrebbe pensare ad acquisire almeno il lemmario del GRADIT⁴⁵, ma non molto di più perché la struttura si è venuta a configurare come chiaramente incompatibile; e lo stesso vale, all'interno delle voci, per altri auspicabili trapianti dal GDLI, la cui automazione appare impossibile.

La scelta, ovviamente, dipende da un cumulo di condizioni che trascendono il singolo gruppo di ricerca, ma qualcosa andrà indubbiamente escogitato: l'unico *caveat* è che il ricorso a questa procedura debba rispondere a condizioni il meno soggettive possibili, ed essere preferibilmente in gran parte automatizzabile.

7. Conclusioni

Abbiamo cercato, muovendoci tra corpora e macrostruttura delle voci, di portare qualche suggerimento, sperabilmente utile, per la stesura del VoDIM.

In particolare, prendendo come punto di partenza il modello tipo-TLIO, che è stato autorevolmente avanzato, ne abbiamo messo in luce alcune problematicità, proponendo possibili strategie per superarle. Nel che fare abbiamo pure constatato come una procedura puramente *corpus driven*, che altrove (cfr. Barbera 2013, pp. 15-18) avevamo mostrato teoricamente non essere ottimale, non sia in questo specifico caso neppure in pratica molto realizzabile: in ciò confermando le presupposizioni iniziali. Questi materiali “extra-corpora” potrebbero provenire da varie fonti, che possono andare dalla Wikipedia al Battaglia. Anzi, tra le opzioni che recentemente si sono affacciate, vista la diponibilità materiale e legale di tali materiali in Crusca, è che al GDLI potrebbe essere assegnato un ruolo più centrale nella costituzione del VoDIM; a questo vi vedo almeno due tipi di obiezioni: dal punto di vista della articolazione interna delle voci, la struttura altamente personale, diciamo così, e sempre mutevole degli articoli del Battaglia è tale da vanificare ogni strategia di importazione automatica; dal punto di vista della scelta del lemmario, oltre che anche qui si riversano le peculiarità del Battaglia (si pensi già solo al modo misto di maneggiare POS e semantica), è determinante soprattutto l'inopportunità, a mio parere almeno, di fare del VoDIM una sorta di clone del GDLI, di cui è già comunque pianificata una sua messa a disposizione online. Che i due progetti dialoghino profittevolmente tra di loro (anche, poi, con link ipertestuali a vicenda) è bene, che l'uno si appiattisca sull'altro, meno.

8. Appendice

Per rendere più perspicua la nostra proposta, ho provato ad illustrare alla moda delle spiegazioni metalessicografiche di solito premesse ai dizionari, la voce (in quanto tra quelle qui trattate è la più articolata) di *alienato* (nome ed aggettivo) in una ipotetica (tra le tante espansioni possibili) dimensione “media”:

44 Soprattutto considerando che il suo attuale baricentro è nell'Otto-Novecento e non nella lingua contemporanea, si può senz'altro migliorarlo ma ben difficilmente renderlo l'unica fonte necessaria.

45 Per la questione, in sé imbarazzante, del copyright, penso che alla Crusca non manchino certo frecce nella faretra.

Alienato, s.m. 'malato di mente, pazzo'.
 Dal participio passato di *alienare*.
alienabile, alienabilità, alienante, alienatamente, alienatorio, alienato, alienato₁, alienare, alienarsi₁, alienarsi₂, alienazione, alienista.

1 'pazzo'. SIN *matto*, ANT *sano di mente*
In questo mondo di distratti, di egocentrici, di saccenti, di noiosi, di indifferenti, di alienati, un buon ascoltatore è un esemplare raro, ricercato e gradito, assai più del conversatore indefesso. (Gasperini, Brunella - Il galateo).
 [21 ess.]

Alienato, a. 'dissociato, ceduto, deprivato'
 Dal participio passato di *alienare*.
alienabile, alienabilità, alienante, alienatamente, alienatorio, alienato, alienato₁, alienare, alienarsi₁, alienarsi₂, alienazione, alienista.

1 'inautentico, reso dissociato dalla realtà' (sociologico e filosofico).
 SIN *falso*, ANT *vero*
Nel «Supermarket», invece, (che è opera più recente e che riflette un ordine di esperienze più «alienato») lo schema astratto chiude in una sorta di favola cattiva i carrelli della abbondanza, le ruote che una goccia di nero rende per sempre immobili, i triangoli e i rettangoli delle merci perfette, senza sapore e odore, miraggi e miti della epoca nostra. (Personaggi e vicende dell'arte moderna).
 [4 ess.]

2 'pazzo'. SIN *matto*, ANT *sano di mente*
Così l'alienista alienato terminò con il peggiore morbo in una casa cosiddetta di salute. (Faldella, Giovanni - Donna Folgore).
 [2 ess.]

3 'ceduto' (giuridico). SIN *venduto*, ANT *mantenuto*. [+11?]
La novità dell'Istituto concerne il diritto del creditore munito di titolo esecutivo di espropriare, senza dover esperire c.c., il bene immobile o mobile registrato alienato a titolo gratuito dal debitore ad un terzo oppure fatto oggetto di un vincolo di destinazione. (AA.VV. - Anno 2016).
 [1 es.]

4 'deprivato'. SIN *privo di*, ANT *dotato di*. [+1?]
la forza poetica e morale di gran parte delle opere presentate» (è ben diverso dal) «generale tono urlante della pittura di oggi, così irrazionalmente tesi al successo, da sostituire al contenuto della vita moderna un triste contenuto alienato di ambizione di ufficialità (7) e di pseudo originalità a tutti i costi. (Personaggi e vicende dell'arte moderna).
 [1 es.]

POS. Indicazione della categoria grammaticale.

Arcidefinizione. Fonte da definire, nell'esempio è mia.

Etimologia. È da decidere se ed in quale voce vada messa.

Derivazione. Quali che siano le decisioni prese per l'etimologia, i rapporti derivazionali vanno sempre esplicitati. Qui ho sperimentalmente fuso i campi 0.0 e 0.2

Link. Tutte le voci in blu nella tavola dovrebbero ancorare dei collegamenti ipertestuali.

Numerazioni. Per migliore leggibilità, nella visualizzazione la numerazione delle sottovoci andrebbe mantenuta, ma quella delle distinzioni sotto lo zero andrebbe omessa (ma non così nelle schede dei redattori, dove la sua presenza anzi garantisce uniformità).

Prime attestazioni. Se fosse possibile metterle, le dovremmo mettere qui, suddivise per ogni sottovoce. Purtroppo non possono essere derivate neppure dal Battaglia, sia perché vengono citate solo le opere e non le date (che sono pertanto ricostruibili solo in modo non automatico), sia anche perché il lemmario e l'articolazione delle voci, altamente idiosincratici, non sono facilmente rimappabili.

Sottovoci. Partizioni semantiche; la fonte delle definizioni da definire, nell'esempio sono mie.

Sinonimi. La fonte dei sinonimi e degli antonimi andrebbe definita; nell'esempio è mia.

Esempi. Se non altrimenti menzionata, si ha da intendere che l'esempio è tratto dal corpus di base.

Frequenze. Numero di attestazioni almeno nel corpus principale; se anche in corpora ausiliari, ne andrebbe preventivamente stabilita una lista.

Registri e domini. Ne andrebbero stabilite due liste distinte, per ottenerne unitarietà ed organicità.

Esempi dubbi. Andrebbe sempre indicato il numero di esempi dubbi che non sono stati conteggiati; cfr. quanto dicevamo prima per la distinzione aggettivo : participio passato.

Bibliografia

- Aa.Vv., *Le Trésor de la langue française (TLF)*, 16 voll. + 1 supplément, Paris, ATILF, 1971-1994; la versione informatizzata, *Le Trésor de la langue française informatisé (TLFi)*, è liberamente disponibile online: <http://stella.atilf.fr/dendien/scripts/tlfiv4/showps.exe?p=combi.htm;java=no>; ¶ TLF: I. A-Af, 1971; II. Af-An, 1973; III. An-Ba, 1974; IV. Ba-Ca, 1975; V. Ca-Co, 1977; VI. Co-Dé, 1978; VII. Dé-Ép, 1979; VIII. Ép-Fu, 1980; IX. G-In, 1981; X. In-Lo, 1983; XI. Lo-Na; XII. Na-Pé, 1986; XIII. Pé-Pt, 1988; XIV. Pt-Sa, 1990; XV. Sa-Te, 1992; XVI. Te-Zz, 1994.
- Aa.Vv., *Eesti kirjakeele seletus-sõnaraamat*, VII köited, [Tallinn], Eesti teaduste akadeemia – Keele ja kirjanduse instituut (fino al IV. köide, 4. vihik), poi Eesti keele instituut, 1988-2007. ¶ I.k A-J: 1.v A-ba, 1988/1989; 2.v ba-fo, 1991; 3.v fo-hü, 1991; 4.v hü-jüü, 1991; II.k K: 1.v K-kee, 1992; 2.v kee-ko, 1992; 3.v ko-ku, 1992; 4.v ku-küü, 1993; III.k L-N: 1.v L-loo, 1992; 2.v loo-me, 1993; 3.v me-mü, 1994; 4.v mü-nüü, 1994; IV.k O-R: 1.v O-pe, 1994; 2.v pe-pl, 1995; 3.v pl-põ, 1996; 4.v põ-ra, 1996; V.k Ra-Ž: 1.v ra-rä, 1997; 2.v rä-se, 1998; 3.v se-su, 1999; 4.v su-žü, 2000; VI.k T-V: 1.v T-ti, 2001; 2.v ti-tu, 2002; 3.v tu-un, 2003; 4.v un-va, 2004; VII.k V-Y: 1.v va-vu, 2005; 2.v vu-yu, 2007.
- Adriano Allora - Carla Marellò, “Ricarica clima”. Accorciamenti nella lingua dei newsgroup, in Cresti E., editor, *Atti del IX Congresso della Società Internazionale di Linguistica e Filologia Italiana (SILFI): “Prospettive nello studio del lessico italiano”* (Firenze, 14-17 giugno 2006), a cura di Emanuela Cresti, Firenze, Franco Cesati, vol. II, pp. 533-538.
- Manuel Barbera, “Partes Orationis”, “Parts of Speech”, “Tagset” e dintorni. Un prospetto storico-linguistico, in *Anatolistica Indoeuropeistica e Oltre – nelle Memorie dei Seminari offerti da Onofrio Carruba (Anni 1997-2002), al Medesimo presentate*, a cura di Guido Borghi ed Alfredo Rizza, Milano, Qu.A.S.A.R., 2011 “Antiqui Aevi grammaticae artis studiorum consensus. Series maior” I, tomo I, pp. 113-145.
- Manuel Barbera, *Per la storia di un gruppo di ricerca. Tra bmanuel.org e corpora.unito.it*, in *Corpora e linguistica in Rete*, a cura di Manuel Barbera, Elisa Corino e Cristina Onesti, Perugia, Guerra Edizioni, 2007 “L'officina della lingua. Strumenti” I, pp. 3-20.
- Manuel Barbera, *Schema e storia del “Corpus Taurinense”. Linguistica dei corpora dell'italiano antico*, Alessandria, Edizioni dell'Orso, 2009. Nuova versione completamente rifatta ed ampliata di *Linguistica dei «corpora» per l'italiano antico. Annotazione morfosintattica di testi fiorentini del Duecento*, a cura di Manuel Barbera e Carla Marellò, Alessandria, Edizioni dell'Orso, 2001.
- Manuel Barbera, *Una introduzione ai NUNC: storia della creazione di un corpus*, in Manuel Barbera, *Molti occhi sono meglio di uno: saggi di linguistica generale 2008-12*, [Milano], Qu.A.S.A.R., 2013, pp. 97-114; anche in francese *Une introduction au NUNC: histoire de la création d'un corpus*, in *Variétés syntaxiques dans la variété des textes online en italien: aspects micro- et macrostructuraux*, édité par Angela Ferrari et Letizia Lala, Nancy, Université de Nancy II, 2011 = «Verbum» XXXIII (2011)¹⁻² 9-36.
- Manuel Barbera, *Linguistica dei corpora e linguistica dei corpora italiana. Un'introduzione*, Milano, Qu.A.S.A.R., 2013. Ebook disponibile a <http://www.bmanuel.org/man/cl-HOME.htm>.
- Manuel Barbera, *Anglicismi e lingua italiana: teoria e prassi*, in *Quanto più la relazione è bella. Saggi di storia della lingua italiana 1999-2014*, Torino - Tricase (LE), bmanuel.org - Youcanprint Self-Publishing, 2015 “bmanuel.org glottologica et philologica - series maior” I, pp. 139-156.
- Manuel Barbera, *Wanderworten etymological format: a ‘hemp’ case study, starting from Votic and Baltofinnic*, in Tinatin Margalidze - George Meladze, *Lexicography and Linguistic Diversity. Proceedings of the XVII EURALEX International Congress, 6 - 10 September 2016, Tbilisi (Georgia): Lexicography and Linguistic Diversity*, Tbilisi, Ivane Javakhishvili Tbilisi State University Lexicographic Centre, 2016, pp. 574-582.

- Manuel Barbera, *Corpora Torinesi per il VoDIM*, poster alla Giornata di Ateneo “Humanities in a day – La ricerca umanistica si racconta”, Torino, UniTO, 13 giugno 2018.
- Manuel Barbera - Simona Colombo, *Pre-processing Normalization Procedures for Newsgroup Corpora*, in *Bootstrapping Information from Corpora in a Cross-linguistic Perspective*, edited by Massimo Moneglia and Alessandro Panunzi, Firenze, Firenze University Press, 2010 “Strumenti per la didattica e la ricerca” 96, pp. 175-191.
- Manuel Barbera - Carla Marello, *Tra scritto-parlato, Umgangssprache e comunicazione in rete: i corpora NUNC*, in «Studi di Grammatica Italiana» XXVII (2008, recte 2011) = *Per Giovanni Nencioni. Convegno Internazionale di Studi*. Pisa - Firenze, 4-5 Maggio 2009, a cura di Anna Antonini e Stefania Stefanelli, Firenze, Le Lettere, 2011, pp. 157-186. Poi raccolto, aggiornato, in Manuel Barbera, *Quanto più la relazione è bella: saggi di storia della lingua italiana 1999-2014*, bmanuel.org - Youcanprint, 2015 “bmanuel.org glottologica et philologica - series maior” 1, pp. 157-182.
- Manuel Barbera - Carla Marello, *Newsgroup e lessicografia: dai NUNC al VoDIM*, in *JADT '18, Proceedings 14th International Conference on the Statistical Analysis of Textual Data (Rome 12-15 June 2018)*, Vol I, Roma, Universitalia, 2018, pp. 76-82.
- Salvatore Battaglia [† 1971 - Giorgio Barberi Squarotti dal vol. IV], *Grande dizionario della lingua italiana*, Torino, U.T.E.T., 1961-.... ¶ I. A-Balb, 1961; II. Balc-Cerr, 1962; III. Cert-Dag, 1964; IV. Dah-Duu, 1966; V. E-Fin, 1968; VI. Fio-Grau, 1970; VII. Grav-Ing, 1972; VIII. Ini-Libb, 1975; IX. Libe-Med, 1975; X. Mee-Moti, 1978; XI. Moto-Orac, 1981; XII. Orad-Pere, 1984; XIII. Perf-Po, 1986; XIV. Pra-Py, 1988; XV. QRia, 1990; XVI. Rib-Roba, 1992; XVII. Robb-Schi, 1994; XVIII. Scho-Sik, 1996; XIX. Sil-Sque, 1998; XX. Squi-Tog, 2000; XXI. Toi-Z, 2002; *Supplemento* a cura di Edoardo Sanguineti, 2004; *Supplemento* a cura di Edoardo Sanguineti, 2009; *Indice degli autori citati* a cura di Giovanni Ronco, 2004.
- Carlo Battisti - Giovanni Alessio, *Dizionario etimologico italiano*, Firenze, G. Barbèra Editore, 1950-1957. ¶ I. A-Ca, 1950, pp. xxxj + 1-820; II. Ca-Fa, 1951, pp. 821-1604; III. Fa-Me, 1952, pp. 1605-2404; IV. Me-Ra, 1954, pp. 2405-3188; V. Ra-Zu, 1957, pp. 3189-4132.
- Tesoro della lingua italiana delle origini*, diretto da Pietro Beltrami, Firenze, CNR - Centro di studi Opera del Vocabolario Italiano, 1998-...; disponibile su <http://tlio.ovi.cnr.it/TLIO/>.
- A Magyar nyelv történeti-etimológiai szótára*, főszerkesztő Benkő Loránd. Szerkesztők: Kiss Lajos és Papp László (I-II. kötet), Kubinyi László és Papp László (III. kötet), S. Hámori Antónia és Zaicz Gábor (IV. kötet), Budapest, Akadémiai Kiadó, 1976-1984. ¶ I. k. 1976, II. k. 1970, III. k. 1976, IV. k. 1984.
- COBUILD => Sinclair 1987.
- Manlio Cortelazzo - Paolo Zolli, *Dizionario etimologico della lingua italiana*, Bologna, Zanichelli, 1979-1988. ¶ I. A-C, 1979, pp. xxvij + 1-307; II. D-H, 1980, pp. xix + 309-536; III. I-N, 1983, pp. xxj + 537-815; IV. O-R, 1985, pp. xxj + 817-1113; V. S-Z, 1988, pp. xx + 1115-1470.
- Mauro Costantino - Carla Marello - Cristina Onesti, *La cucina discussa in rete. Analisi di gruppi di discussione italiani relativi alla cucina*, in *Atti del convegno ASLI 2007 “Storia della lingua e storia della cucina. Parola e cibo: due linguaggi per la storia della società italiana”*. Modena, 20-22 settembre 2007, a cura di Cecilia Robustelli e Giovanna Frosini, Firenze, Franco Cesati, pp. 717-727.
- DEI = Battisti - Alessio 1950-57.
- DELI = Cortelazzo - Zolli 1979-88.
- Grande dizionario italiano dell'uso*, ideato e diretto da Tullio De Mauro, con la collaborazione di Giulio C. Lepschy e Edoardo Sanguineti, Torino, U.T.E.T., 1999, voll. 6. Con aggiornamenti: *Nuove parole italiane dell'uso*, 2003; *Nuove parole italiane dell'uso*. 2., 2007.

- [Giacomo Devoto - Gian Carlo Oli], *Nuovo Devoto-Oli. Il vocabolario dell'italiano contemporaneo 2019. Con App scaricabile su smartphone e tablet*, a cura di Luca Serianni, Firenze, Le Monnier, 2018.
- DISC = Sabatini - Coletti 2008.
- EKS = Aa.Vv. 1988-2007.
- GDLI = Battaglia *et alii* 1961-2009.
- GRADIT = De Mauro 1999.
- Biblia sacra iuxta vulgatam versionem*, adiuvantibus B[onifatio] Fischer, I[ohanne] Gribomont, H[aldley] F[rederick] D[avis] Sparks, W[alter] Thiele, recensuit et breve apparatus critico instruxit Robert Weber, editionem quintam emendatam retractatam preparavit Roger Gryson, Stuttgart, Deutsche Bibelgesellschaft, 2007 [1969].
- Riccardo Gualdo, *Un nuovo vocabolario dinamico dell'italiano: il lessico specialistico e settoriale*, in «Studi di lessicografia italiana» XXXV (2018) 193-216.
- Adam Kilgarriff, *Googleology is Bad Science*, in «Computational Linguistics» XXXIII (2007)¹ 147-151.
- Kalevi Kull, *Jakob von Uexküll: an Introduction*, in «Semiotica» CXXXIV 134 (2001)^{1/4} 1-59.
- Patrizia Bertini Malgarini - Marco Biffi - Ugo Vignuzzi, *Dal corpus al dizionario: prime riflessioni lessicografiche sul Vocabolario storico della cucina italiana postunitaria (VoSCIP)*, in *JADT '18, Proceedings 14th International Conference on the Statistical Analysis of Textual Data (Rome 12-15 June 2018)*, Vol I, Roma, Universitalia, 2018, pp. 90-98.
- Carla Marellò, *Does Newsgroups "Quoting" Kill or Enhance Other Types of Anaphors?*, in *Comparing Anaphors between Sentences, Texts and Languages*, edited by Iørn Korzen and Lita Lundquist, København, Samfundslitteratur Press, 2007, pp. 145-157.
- Dario Martinelli, *A Critical Companion to Zoosemiotics: People, Paths, Ideas*, Dordrecht - Heidelberg - London - New York, Springer, 2010 "Biosemiotics" 5.
- W[ilhelm] Meyer-Lübke, *Romanisches Etymologisches Wörterbuch*, Heidelberg, Carl Winter Universitätsverlag, 1972⁵ [=1935₃] "Sammlung Romanischer Elementar- und Handbücher. Dritte Reihe: Wörterbücher".
- Inga Pollmann, *Invisible Worlds, Visible: Uexküll's Umwelt, Film, and Film Theory*, in «Critical Inquiry» XXXIX (2013)⁴ 777-816.
- REW = Meyer-Lübke 1935.
- Il Sabatini Coletti. Dizionario della lingua italiana*, diretto da Francesco Sabatini e Vittorio Coletti, Milano, Rizzoli Larousse, 2008, 1 vol. + 1 CD-ROM.
- Max Scheler, *Jakob Baron von Uexküll: Bausteine zu einer biologischen Weltaanschauung*, Besprechung, in «Beilage zu Die weißen Blätter» I (1914)⁶ 119-121.
- Thomas A. Sebeok, *Review of Communication among social bees by Martin Lindauer; Porpoises and Sonar by Winthrop N. Kellogg; Man and Dolphin by John C. Lilly*, in «Language» XXXIX (1963)³ 448-466.
- Salvatore Claudio Sgroi, *L'inglese e l'italiano: alternativa? No. Arricchimento: sì*, in *Lingua italiana*, Istituto della Enciclopedia Italiana, 12 luglio 2018. Voce online: http://www.treccani.it/magazine/lingua_italiana/articoli/scritto_e_parlato/Inglese.html.
- Collins COBUILD – English Language Dictionary*, editor in chief John [Mc Hardy] Sinclair, London, Harper Collins, 1987.
- TESz = Benkő *et alii* 1976-84.
- TLF = Aa.Vv. 1971-94.
- TLFi = Aa.Vv. 1971-94.
- TLIO = Beltrami 1998-...

- [N(icolò) Tommaseo], *Nuovo dizionario de' sinonimi della lingua italiana*, Firenze, dalla Tipografia di Luigi Pezzati, 1830.
- Dizionario della lingua italiana*, nuovamente compilato dai signori Nicolò Tommaseo e cav. professore Bernardo Bellini, con oltre 100.000 giunte ai precedenti dizionarii, Torino [ma II.1-IV.1, Torino - Napoli; e IV.2, Pisa - Livorno - Napoli], dalla Società L'Unione Tipografico-Editrice, 1865-1879.
- ¶ Vol. I.1 A-AZZ, 1865; Vol. I.2 B-CZA, 1865; Vol. II.1 D-FUT, 1869; Vol. II.2 G-LUV, 1869; Vol. III. M-QUO, 1871; Vol. IV.1 R-SVO, 1872; Vol. IV.2 T-ZUZ, *Giunte e correzioni, Tavola delle abbreviazioni*, 1879.
- J[akob Johann baron] von Uexküll, *Umwelt und Innenwelt der Tiere*, zweite vermehrte und verbesserte Auflage, Berlin - Heidelberg, Springer Verlag, 1921.
- Luca Valle, *Varietà diafasiche e forestierismi nell'italiano nei gruppi di discussione in rete*, in *Perspectives fonctionnelles: emprunts, économie et variations dans les langues. S.I.L.F. 2004. XXVIII Colloque de la Société internationale de linguistique fonctionnelle, tenu à Saint-Jacque-de-Compostelle et à Lugo du 20 au 26 septembre 2004*, édité par Montserrat López Díaz et Maria Montes López, Lugo, Editorial Axac, 2006, pp. 371-374.
- Ugo Vignuzzi - Patrizia Bertini Malgarini, *Bagnomaria nel Vocabolario storico della cucina italiana postunitaria (VoSCIP)*, in *Di tutti i colori. Studi linguistici per Maria Grossmann*, a cura di Roberta D'Alessandro, Gabriele Iannàccaro, Diana Passino ed Anna M. Thornton, [Utrecht], Utrecht University, 2017, pp. 369-384.
- Lo Zingarelli 2017. *Vocabolario della lingua italiana*. A cura di Mario Cannella e Beata Lazzarini, Bologna, Zanichelli. 2017.