

STRUMENTI  
PER LA DIDATTICA E LA RICERCA

– 109 –

UNIVERSITÀ DEGLI STUDI DI FIRENZE  
DIPARTIMENTO DI ITALIANISTICA

BIBLIOTECA DIGITALE

COMITATO SCIENTIFICO

Adele Dei  
Anna Dolfi  
Simone Magherini  
Massimo Moneglia

Volumi pubblicati:

MODERNA [diretta da Anna Dolfi]

1. *Giuseppe Dessì. Storia e catalogo di un archivio*, a cura di Agnese Landini, 2002.
2. *Le corrispondenze familiari nell'archivio Dessì*, a cura di Chiara Andrei, 2003.
3. Nives Trentini, *Lettere dalla Spagna. Sugli epistolari a Oreste Macrì*, 2004.
4. *Lettere a Ruggero Jacobbi. Regesto di un fondo inedito con un'appendice di lettere*, a cura di Francesca Bartolini, 2006.
5. «L'Approdo». *Copioni, lettere, indici*, a cura di Michela Baldini, Teresa Spignoli e del GRAP, sotto la direzione di Anna Dolfi, 2007 (CD-Rom allegato con gli indici della rivista e la schedatura completa di copioni e lettere).
6. Anna Dolfi, *Percorsi di macritica*, 2007 (CD-Rom allegato con il *Catalogo della Biblioteca di Oreste Macrì*).
7. *Ruggero Jacobbi alla radio*, a cura di Eleonora Pancani, 2007.
8. Ruggero Jacobbi, *Prose e racconti. Inediti e rari*, a cura di Silvia Fantacci, 2007.
9. Luciano Curreri, *La consegna dei testimoni tra letteratura e critica. A partire da Nerval, Valéry, Foscolo, d'Annunzio*, 2009.
10. Ruggero Jacobbi, *Faulkner ed Hemingway. Due nobel americani*, a cura di Nicola Turi, 2009.
11. Sandro Piazzesi, *Girolamo Borsieri. Un colto poligrafo del Seicento. Con un inedito «Il Salterio Affetti Spirituali»*, 2009.
12. *A Giuseppe Dessì. Lettere di amici e lettori. Con un'appendice di lettere inedite*, a cura di Francesca Nencioni, 2009.
13. Giuseppe Dessì, *Diari 1949-1951*, a cura di Franca Linari, 2009.

LINGUISTICA [diretta da Massimo Moneglia]

1. *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*, edited by Massimo Moneglia and Alessandro Panunzi, 2010.
2. Alessandro Panunzi, *La variazione semantica del verbo essere nell'italiano parlato*, 2010.
3. *Language, Cognition and Identity. Extensions of the endocentric/exocentric language typology*, edited by Emanuela Cresti and Iørn Korzen, 2010.

INFORMATICA E LETTERATURA [diretta da Simone Magherini]

1. *BIL Bibliografia Informatizzata Leopardiana 1815-1999. Manuale d'uso vers. 1.0*, a cura di Simone Magherini, 2003.

# **Language, Cognition and Identity**

Extensions of the endocentric/exocentric  
language typology

*edited by*  
Emanuela Cresti  
Iørn Korzen

Firenze University Press  
2010

Language, Cognition and Identity : Extensions of the endocentric/exocentric language typology / edited by Emanuela Cresti and Iørn Korzen. – Firenze : Firenze University Press, 2010.

(Strumenti per la didattica e la ricerca ; 109)

<http://digital.casalini.it/9788864532264>

ISBN 978-88-6453-220-2 (print)

ISBN 978-88-6453-226-4 (online)

Immagine di copertina: Alessandro Geri Rustighi

Progetto grafico di Alberto Pizarro Fernández

© 2010 Firenze University Press  
Università degli Studi di Firenze  
Firenze University Press  
Borgo Albizi, 28, 50122 Firenze, Italy  
<http://www.fupress.com/>

*Printed in Italy*

# INDEX

INTRODUCTION	VII
PREDICATI E SOSTANTIVI COMPLESSI: COMPLEMENTARITÀ E ISOMORFIA <i>Michael Herslund</i>	1
KANTIAN GRAMMAR APPLIED TO FRENCH, DANISH AND ENGLISH <i>Hanne Korzen</i>	9
I VERBI GENERALI NEI CORPORA DI PARLATO. UN PROGETTO DI ANNOTAZIONE SEMANTICA CROSS-LINGUISTICA <i>Massimo Moneglia, Alessandro Panunzi</i>	27
ITALIANO LINGUA ESOCENTRICA: L'USO DEI TEMPI VERBALI NELLA NARRAZIONE <i>Luisa Amenta</i>	47
DA UNA LINGUA ENDOCENTRICA AD UNA LINGUA ESOCENTRICA? IL CASO DELL'ITALIANO <i>Emanuela Cresti</i>	61
L'ITALIANO IN UNA PROSPETTIVA DI TREEBANK. IL COPENHAGEN DEPENDENCY TREEBANK PROJECT: ASPETTI SINTATTICI E TESTUALI <i>Iørn Korzen</i>	77
SPANISH IN A TREEBANK PERSPECTIVE. THE COPENHAGEN DEPENDENCY TREEBANK PROJECT. MORPHOLOGICAL ASPECTS <i>Henrik Høeg Müller</i>	99
LE RISORSE DI RETE DELL'ITALIANO. PRESENTAZIONE DEL PROGETTO "RIDIRE.IT" <i>Massimo Moneglia, Samuele Paladini</i>	111
ANCORA SULLE "RISORGENZE DIALETTALI". PALERMO E IL SUO REPERTORIO FRA CONTINUITÀ E INNOVAZIONE <i>Mari D'Agostino</i>	129

DYSPHEMISM AND NATIONAL, REGIONAL AND CLASS IDENTITY IN CONTEMPORARY ITALIAN <i>Nora Galli de' Paratesi</i>	141
LINGUA E IDENTITÀ. LA POLITICA NELLA RETE DI FACEBOOK <i>Marco Gargiulo</i>	155
HUMOUR IN INTERCULTURAL PROFESSIONAL SETTINGS. A SHORTCUT TO LANGUAGE, COGNITION AND IDENTITY <i>Lita Lundquist</i>	167
LEXICAL SEMANTICS NEGOTIATED. ASSESSING THE FAIRNESS AND MISLEADING POTENTIAL OF ESTABLISHED AND NOVEL FOOD NAMES <i>Viktor Smith</i>	191

## INTRODUCTION

This volume contains a selection of papers presented at the two day Italian-Danish linguistic seminar entitled *Lingua, cognizione e identità: estensioni della tipologia delle lingue endo- ed esocentriche*. The seminar was held at the Italian Department of the University of Florence on the 22<sup>nd</sup> and 23<sup>rd</sup> of September 2009 and was co-organised by Emanuela Cresti and Massimo Moneglia, Florence, and Iørn Korzen, Copenhagen Business School (CBS), as a part of the LABLITA workshop in Corpus Linguistics. Korzen is a member of the CBS based research group “TypoLex”, which for the past 13 years has been exploring linguistic typology, focussing mainly on the typological differences between the Germanic and Romance languages, which the group refers to as endocentric and exocentric languages, respectively. The scholars are led by Michael Herslund and have organised annual conferences and seminars at various European universities. The seminar in Florence was the group’s eleventh seminar in general and its fourth in Italy, following similar events in Turin (1999), Rome (Roma Tre, 2003) and Cagliari (2007).

According to the group’s terminology, an endocentric language is a language whose verbs are lexically precise and concrete whereas its nouns are abstract and vague. Therefore, the informational weight can be said to be located in the centre of the sentence (hence “endo-” [within] “-centric”). On the other hand, an exocentric language has lexically precise and concrete nouns and abstract verbs, thus locating the informational weight outside the verbal centre of the sentence (hence “exo-” [outside] “-centric”). The Germanic languages prove to be (more or less clear-cut) endocentric languages and the Romance languages exocentric, and the mentioned lexical differences entail differences at other levels as well, linguistic (e.g. in the textual structure and in the choice of anaphors) as well as extralinguistic (e.g. in dimensions such as cognition and identity).

The papers presented at the seminar and now compiled in this multilingual volume deal in various ways with these phenomena and are grouped into three main parts. The first part contains contributions focused primarily on grammar and addressing the endo-/exocentric approach theoretically with examples from the Romance languages Italian and French and the Germanic languages English, German and Danish. In *Predicati e sostantivi complessi: complementarità e isomorfia*, Michael Herslund (CBS) investigates the extent to which the basic

complementarity between the exocentric and endocentric languages is repeated in complex expressions, or secondary lexicalization. In exocentric languages, the canonical pattern is one of abstract verbs expressing e.g. PATH and concrete nouns expressing CONFIGURATION, whereas the opposite is the case in endocentric languages, where verbs express MANNER and nouns FUNCTION. When creating concrete verbal expressions of MANNER in exocentric languages, a procedure of compounding is chosen (viz. *venire correndo*), whereas in endocentric languages the secondary lexicalization creating more abstract verbal expressions has the shape of a derivation (viz. *Hinauslaufen*). When creating abstract nouns expressing FUNCTION rather than CONFIGURATION, exocentric languages use derivation (viz. *sale – saliera*), in contrast to endocentric languages, which in order to create nouns expressing CONFIGURATION rather than FUNCTION use compounding (viz. *Salz – Salznapf, Salzstreuer*). The basic complementarity of the primary lexicalization is thus repeated in the creation of expressions taking on the features that are characteristic of the opposite type.

In her paper *Kantian grammar applied to French, Danish and English*, Hanne Korzen (CBS) continues her investigation of Kant, whom - according to Korzen - many linguists refer to, though seemingly without taking him seriously. The scholar shows that a closer look at Kant's cognitive model yields insight into important aspects of the syntactic-semantic constitution of the sentence in different languages. Her main study object is the difference between space and time adjuncts on the one hand, and causal adjuncts on the other. It appears that causal adjuncts behave differently from adjuncts of place and time and that their behaviour and the way they differ from time and place adjuncts is the same in French, Danish and English although the three languages are radically different in many other respects. Hanne Korzen has created a sentence model that seems to fit nicely into Kant's cognitive model, and which may indicate that we are dealing with universal phenomena.

The paper *I verbi generali nei corpora di parlato. Un progetto di annotazione semantica cross-linguistica* by Massimo Moneglia and Alessandro Panunzi focuses on one main lexical correlation of exocentric languages. General verbs are action-oriented predicates which, in their own lexical meaning, refer to many different action types. Activity verbs, on the contrary, are very specific and refer to one only action type. Thus the verbal lexicon shows a typological division of action concepts. Languages may differ with respect to the distribution of verbs between the two classes. English and the Romance languages show a strong preference for general verbs in oral communication. The paper focuses on the fact that the action types extended by general verbs in different languages are not in a one-to-one relationship and that their actual range of variation is presently unknown. For this reason the translation of action verbs is strongly underdetermined. The paper claims, however, that cross-linguistic correspondences are in principle predictable, since the translation of the corresponding English, French and Spanish verbs turns



out to be productive in all instances of each type (once types are identified). This result constitutes the basis for the IMAGACT project, which will build a translation infrastructure for action concepts bootstrapping information from spontaneous speech corpora.

The typological distinction between endocentric and exocentric languages is grounded on a different lexical distribution, which at the same time correlates with macro-structural properties of the language. For instance in I. Korzen's model (I. Korzen 2002 & 2004) endocentric languages are claimed more "paratactic" while exocentric languages, like Italian, are basically "hypotactic". The article by Luisa Amenta *Italiano lingua esocentrica: l'uso dei tempi verbali nella narrazione* exploits a corpus of narratives taken from the *Atlante linguistico della Sicilia* (ALS), and focuses on how speakers from different age groups and educational backgrounds apply verbal tenses in narratives. Amenta shows strong diversity according to diastatic variation regarding the use of tenses. More specifically, the exocentric tendency of Italian is confirmed for those speakers who are more familiar with the standard use of Italian. Such informants are able to use both perfective and imperfective tenses, thus marking the focal relations among events.

From the point of view of historical linguistics, the derivation of Italian from Latin is considered a change from an endocentric to an exocentric language. The most relevant features of this structural modification at the various linguistic levels (phonetics, morphology, syntax) are summarized in the paper *Da una lingua endocentrica ad una esocentrica? Il caso dell'italiano*. Emanuela Cresti claims that the emergence of exocentric features in Italian can be understood as a "creolization process". However, when considering the verbal lexicon of Latin and Italian in the light of the endo-/exocentric typology, they show peculiar features that do not correlate strictly with the overall endo-/exocentric distinction. Latin records movement verbs encoding MANNER, but also movement verbs encoding PATH (through prefixes). Italian records a good deal of movement verbs encoding MANNER.

The second section presents a Danish and an Italian project aimed at the creation of linguistic infrastructures that have proved essential to empiric linguistic and cross-linguistic studies, i.e. treebank and web corpora. Iørn Korzen and Henrik Høeg Müller (CBS) both describe the Copenhagen Dependency Treebank project (CDT), which is a project that involves parallel treebanks for the exocentric languages Italian and Spanish and the endocentric languages Danish, English, and German. The corpora are currently being annotated with respect to syntax, discourse, anaphora, morphology, and translational equivalence. Korzen's paper, *L'italiano in una prospettiva di treebank*, starts with a brief description of the CDT syntax annotation and then focuses on the discourse annotation model with examples taken from the Italian part of the CDT corpus. The exocentric languages are characterized

by their relatively more complex and hypotactic text structure, whereas the endocentric languages, especially the Scandinavian ones, tend to follow a more simple and paratactic text structure. This means that an exocentric sentence may very well contain two or more propositions that would correspond to two or more sentences in a Scandinavian language. Therefore, many of the CDT's syntax annotation labels are found also in the discourse annotation.

In his paper, *Spanish in a treebank perspective*, Henrik Høeg Müller focuses on the CDT annotation design for NP morphology. The internal structure of words and word-like phrases is encoded as a dependency tree that can be specified in two different ways: either as an ordinary dependency tree, i.e. similar to the CDT syntactic annotation, or by means of an abstract operator specification. The dependency annotation encodes the structure between word-like phrases, i.e. phrasal compounds and regular NPs, while the operator annotation encodes dependency structure within solid orthography compounds and derivationally constructed words. With respect to NP-structure in a lexical typological perspective, it is claimed that a natural consequence of the semantic vagueness of endocentric (e.g. Danish) simple nouns is that Danish makes use of compounding. On the other hand, exocentric (e.g. Spanish) simple nouns are semantically rich, and therefore exocentric languages are primarily derivational.

Massimo Moneglia and Samuele Paladini present the RIDIRE project (*Risorsa Italiana Dinamica di Rete*, funded by the Italian National Research Program, FIRB), and aimed at compiling, through the use of crawling techniques, a repository of the Italian language that exploits the Italian contents on the Internet. The database will collect massive amounts of freely downloadable documents, covering all the possible domains of language use: law, religion, politics, literature, trade etc. RIDIRE.it is promoted, and will be disseminated and maintained by SILFI (*Società internazionale di linguistica e filologia italiana*), and it is designed for use by all parties involved in the teaching of Italian abroad. RIDIRE.it will be accessible and searchable online, and therefore easy to use for language acquisition purposes thanks to a set of user-friendly tools, which will also ensure valid computation of the linguistic data.

The last section of this volume focuses on the topic "Language and Identity" and presents a number of Italian, French and Danish examples of how socio-cultural reflections of the three languages manifest themselves in various areas of communication: in dialects, in markers of diastatic levels, in social network, in humour and in the creation of specific terminology.

The dynamics of language varieties available to the population in a given region can be ascribed to at least three factors: language use, language representations and institutional interventions. In Italy, all three factors interact. The article by Mari D'Agostino *Ancora sulle "risorgenze dialettali". Palermo e il suo repertorio fra*

*continuità e innovazione* focuses on changes recorded in the language varieties used by teenagers and immigrants in the Palermo area, i.e. the part of the population which, at various levels, is involved in a language learning process. Also in this case, the source of data is the *Atlante linguistico della Sicilia* (ALS). Today, the Palermo area shares with the rest of Sicily the same “quiet language ideology”: The higher the level of education, the more frequent the shifts from Standard Italian to Sicilian in language performance. Nowadays, the dialect is considered a source of cultural richness, rather than a negative mark, and its use is not discouraged by parents.

The link between language style and socio-political identity in Italian is the topic of the articles by Nora Galli de’ Paratesi and Marco Gargiulo. *Dysphemism*, i.e. the intentional use of taboo words, is a term used to indicate the opposite of *euphemism*. Dysphemisms often fill the whole discourse in an irrational manner and are used not for denotation purposes but to convey heavy negativity, aggressiveness, contempt and rage. Nora Galli de’ Paratesi estimates the amount of taboo words in today’s Italy by comparing with a 1964 study on the use of euphemism. The article *Dysphemism and national, regional and class identity in contemporary Italian* determines a vast difference between then and now and argues that there are three reasons for the much more frequent use of taboo words today. One reason is purely linguistic. The weakening of dialects has caused Italian to be used at more familiar levels and thus with the lower degree of formality that used to be reserved for dialects. The two other reasons are socio-political: a bitterly aggressive political discourse that has spilt over into the general discourse and has created a romantic, false nostalgia for the spontaneity of dialects, and finally a violent political scene.

In the Internet 2.0 era the variety of communicative means used for political propaganda is increasing. Many web sites promote politicians through political forums in which citizens interact directly with their leaders. Also on-line newspapers, guest comments and readers’ reactions to political arguments play a role, as can be seen from the early results of research carried out on political communication in social networks, which are now some of the new emerging contexts exploited for political marketing. In his paper *Lingua e identità: la politica nella rete di Facebook* Marco Gargiulo reports on various instances of these types of political communication on Facebook. The author confirms that the use of new media conveys the emergence of oral features in written texts and argues that, when focussing on political arguments, ordinary communication is strongly influenced by the style that has emerged on the Italian political arena over the past years. The texts show a simplified language with a paratactic construction, which is characterized by the use of dialect and a strong aggressive modality (and totally devoid of any euphemistic strategy, in contrast to former usage). These choices mark the identity of the group while causing the link between language and national identity to be lost.

In her paper *Humour in intercultural professional settings. A shortcut to language, cognition and identity*, Lita Lundquist explores the relationship between language, cognition and identity via the study of humour. Within a comparative analysis of how Danes and Frenchmen use and react to humour in professional settings with each other, differences in this respect between the two language communities are explained both in a linguistic framework as fundamental differences between the Danish (endocentric) and French (exocentric) language, and in a sociological framework as a difference in ‘identities’ between the Danish and the French “national character” or “national culture”. This latter difference is explained, in a last step, as being the result of different processes of civilization.

In *Lexical semantics negotiated*, Viktor Smith examines the creation of food names. In this context, the name is not only crucial to the product’s cognitive identity, but also to its legal identity. The paper reports a review of 821 cases on misleading food naming and labelling processed by the Danish food authorities during the period 2002-2009. The pre-theoretical assumptions and arguments put forward by the immediate actors in the case files are transposed into more exact theoretical terms pinpointing the lingua-cognitive essence of the fundamentally different conflict scenarios that emerge from established (familiar) and novel (unfamiliar) food names, respectively. The ultimate goal of Smith’s research is to contribute new insights into the complexity of online word comprehension, lexicalization and acquisition processes. While the primary focus of this article is monolingual, in casu Danish, it also briefly considers the additional complexities that emerge when food naming practices are compared across linguistic and typological borders.

Emanuela Cresti & Iørn Korzen  
December 2010

# PREDICATI E SOSTANTIVI COMPLESSI: COMPLEMENTARITÀ E ISOMORFIA

Michael Herslund

Copenhagen Business School

## 1. Introduzione

L'ipotesi fondamentale della nostra teoria di tipologia lessicale (cfr. per es. Herslund 2003; 2007; 2009) è che i due tipi – lingue esocentriche (romanze) e lingue endocentriche (germaniche) – si distinguono in maniera sistematica nella scelta canonica di componenti semantiche realizzate dai due tipi nella loro lessicalizzazione, cioè nelle combinazioni delle componenti che ciascun tipo permette nei propri verbi e nei propri sostantivi semplici. Il confronto dei due tipi svela una distribuzione delle componenti semantiche di una complementarità fondamentale:

- nel tipo esocentrico (lingue romanze), i verbi sono generali e astratti: è una “idea” verbale, per esempio quella di ‘movimento’ con la sua direzione (per es. italiano *andare, venire, entrare, uscire, salire, scendere* ...). I sostantivi sono specifici e concreti: è un’“immagine” dell’oggetto (per es. *bricco, brocca, cuccuma, caffettiera* ..., che denotano differenti tipi di recipienti per contenere e servire liquidi);
- nel tipo endocentrico (lingue germaniche), i verbi sono specifici e concreti: è un’“immagine” verbale. Per descrivere un movimento bisogna esprimere anche la maniera in cui è eseguito il movimento (corso, nuoto, a piedi ...). I sostantivi sono generali e astratti: è l’“idea” dell’oggetto, cioè la sua funzione, il suo scopo (per es. tedesco *Kanne*, che denota tutti i recipienti destinati a contenere e a servire liquidi).

Si tratta dunque di una distribuzione complementare dove da un lato abbiamo delle componenti semantiche che rendono l’aspetto visibile della situazione verbale o

dell'oggetto descritto dal sostantivo (immagine), e dall'altro delle componenti che rendono l'aspetto "ideale" o "teleologico" (idea):

Tabella 1

	Verbo	Sostantivo
<i>Esocentrico</i>	idea	<i>immagine</i>
<i>Endocentrico</i>	<i>immagine</i>	idea

La visibilità della situazione verbale risulta dalla scelta nella lessicalizzazione verbale della componente [MANIERA] nel tipo endocentrico (per es. ted. *laufen* 'correre'), mentre la visibilità di un oggetto risulta dalla scelta nella lessicalizzazione nominale della componente [CONFIGURAZIONE] nelle lingue esocentriche (per es. it. *bricco*).

## 2. Verbi o predicati complessi

Questa complementarità fra i due tipi si riproduce anche quando si tratta di esprimere la "componente mancante", ossia il concetto dominante nell'altro tipo, cioè la [MANIERA] nel verbo esocentrico o la [DIREZIONE] nel verbo endocentrico, la [FUNZIONE] nel sostantivo egocentrico o la [CONFIGURAZIONE] nel sostantivo endocentrico. Questa lessicalizzazione "secondaria" è necessaria perché una lingua possa esprimere tutto (cfr. Jakobson 1959).

Come abbiamo visto, i verbi del tipo esocentrico lessicalizzano canonicamente la componente [DIREZIONE] nei propri verbi di movimento, ma non quella di [MANIERA] – a pari di numerosi altri verbi (verbi generali, cfr. Moneglia & Panunzi in questo volume).

Per esprimere anche il senso di [MANIERA] in una espressione di movimento nel tipo esocentrico bisogna dunque aggiungere qualche cosa al verbo generale. Quest'operazione ha la forma di una creazione di un predicato complesso per l'aggiunta di un verbo infinito, un gerundio. Questa costruzione è comune a tutte le lingue romanze:

- (1) Fr. Elle sort en courant.  
 It. Veniva correndo verso di me.  
 Sp. Platero llegó trotando.

Dal momento che si tratta di un'operazione che combina due elementi lessicali (due radici verbali), possiamo parlare di un processo di *composizione*: due verbi in uno – un predicato complesso.

Nel tipo endocentrico dall'altro lato, i verbi lessicalizzano canonicamente la [MANIERA], ma rifiutano sistematicamente la componente [DIREZIONE]. Per esprimere anche questo concetto bisogna dunque aggiungere qualche cosa, un satellite secondo la terminologia di Talmy (1985). A tal fine si procede a un'operazione di pre- o suffissazione, in tedesco per esempio:

- (2) Ted. a. Er läuft hinaus.  
 [Esce fuori correndo]  
 b. ... dass er hinausläuft.  
 [... che esce fuori correndo]  
 c. Er will hinauslaufen.  
 [Vuole uscire fuori correndo]

Senza entrare nella discussione del gioco complesso delle manifestazioni del satellite come prefisso o particella postverbale nelle lingue germaniche, o addirittura l'omissione del verbo (*Er will \_\_ hinaus* 'Vuole uscire'), possiamo constatare che si tratta di un processo di *derivazione* combinante un elemento lessicale con un elemento grammaticale: la creazione di un verbo derivato (*hinauslaufen – läuft hinaus*).

Per produrre un verbo che esprima il concetto fondamentale caratteristico del tipo opposto, le lingue esocentriche procedono dunque a un'operazione di composizione, le lingue endocentriche a una di derivazione.

### 3. Sostantivi complessi

Nei sostantivi – vista l'ipotesi della complementarità – i fatti devono essere inversi: il sostantivo esocentrico semplice, cioè non derivato, contiene canonicamente la componente [CONFIGURAZIONE], ma non o di rado quella di [FUNZIONE]. Per ottenere un sostantivo che contenga tale componente, dunque una lessicalizzazione secondaria, si deve aggiungere qualche cosa, e questo qualche cosa è spesso un suffisso abbastanza generale aggiunto a una radice verbale (a) o nominale (b):

- (3a) Fr. *dormir – dortoir*  
*écrire – bureau/écritoire*  
*arroser – arrosoir*  
 It. *dormire – dormitorio*  
*scrivere – scrivania, scrittorio*  
*annaffiare – annaffiatoio*  
 Sp. *dormir – dormitorio*  
*escribir – escribanía, escritorio*  
*regar – regadera*

- (3b) Fr. *encre – encrier*  
*cendre - cendrier*  
*sel – salière*  
*café – cafetière*
- It. *inchiostro – calamaio*  
*cenere – ceneriera/portacenere*  
*sale – saliera*  
*caffè – caffettiera*
- Sp. *tinta – tintero*  
*ceniza – cenicero*  
*sal – salero*  
*café – cafetera*

Come si vede, il procedimento è quello di una derivazione, con certe eccezioni purtroppo come l'uso di una parola non imparentata (fr. 'bureau', it. 'calamaio') o una composizione (it. 'portacenere'). Ma quest'uso segue infatti il modello canonico della lessicalizzazione sostantivale primaria dove, nelle lingue romanze (cfr. Herslund 2000; Baron 2000; 2003; Korzen 2005), si trovano soprattutto serie di parole non imparentate come *macchina, automobile, camion, furgone* (cfr. Korzen 2005: 37), mentre le lingue germaniche (endocentriche) costruiscono tali serie come sostantivi composti (iponimi) con una radice comune (iperonimo) come il tedesco *Wagen* (app. 'veicolo'): *Personwagen, Lastwagen, Güterwagen* (cfr. Herslund 2007: 9).

L'importante è pertanto che il sostantivo così derivato non rivela più la configurazione dell'oggetto – o almeno solo in maniera indiretta – ma menziona soltanto la sua funzione.

Il sostantivo endocentrico è sempre, o al meno in modo predominante, costruito come una composizione, cioè la combinazione di due elementi lessicali, per indicare la funzione dell'oggetto. Corrispondendo alle serie romanze di (3a/b), abbiamo dunque in tedesco la serie (3c) che segue esattamente il modello canonico della lessicalizzazione primaria sottolineante la funzione dell'oggetto:

- (3c) Ted. *schlafen – Schlafsaal*  
*schreiben – Schreibtisch*  
*gießen – Gießkanne*
- Tinte – Tintenfass*  
*Asch – Aschbecher*  
*Salz – Salznapf, Salzstreuer*  
*Kaffee – Kaffeekanne*

Dato che la componente [FUNZIONE] è dominante nel sostantivo endocentrico, bisogna aggiungere qualche cosa per esprimere la [CONFIGURAZIONE], e ciò che



viene aggiunto è un altro elemento lessicale. Si tratta dunque ancora di una composizione:

- (4) Ted. *Kanne – Glas-kanne* (app. ‘bricco di vetro’)

Con tale creazione, il sostantivo endocentrico esprime la configurazione dell’oggetto – è un’immagine – e non menziona che in maniera indiretta la sua funzione esatta (per caffè, té, vino, latte o altri liquidi).

La composizione è, come si è visto poco sopra, il procedimento normale per la formazione di iponimi nelle lingue germaniche (endocentriche), e quell’operazione si ripete anche nell’iponimia “secondaria”:

- |     |                     |                         |                        |
|-----|---------------------|-------------------------|------------------------|
| (5) | Iperonimo:          | <i>Kanne</i>            | app. ‘cuccuma’         |
|     | Iponimo primario:   | <i>Kaffee-kanne</i>     | ‘caffettiera’          |
|     | Iponimo secondario: | <i>Glas-kaffeekanne</i> | ‘caffettiera di vetro’ |

La lessicalizzazione primaria (formazione di iponimi) precisa, come detto, la funzione dell’oggetto: *Kanne – Kaffeekanne* (‘caffettiera’), da cui si può spesso dedurre la sua forma, cfr. Baron (2003). Ma una lessicalizzazione secondaria precisa – oltre alla funzione – anche la configurazione dell’oggetto:

- (6) *Kanne* → *Kaffee-kanne* → *Glas-kaffeekanne*

Un confronto più specifico fra l’italiano e il tedesco sarebbe utile qui: in italiano si deriva per esempio da *sale* la parola *saliera* per ottenere l’espressione di una [FUNZIONE] senza menzionare la configurazione dell’oggetto, mentre in tedesco, per ottenere l’espressione di configurazioni diverse, alla parola italiana corrispondono due parole *Salznapf* e *Salzsteuer*: la prima denota un oggetto aperto, una scodellina per sale, la seconda un oggetto chiuso con piccoli buchi per servire il sale spargendolo (cfr. ted. *streuen* ‘spargere’).

Per produrre un sostantivo che esprima il concetto fondamentale caratteristico del tipo opposto, le lingue esocentriche procedono dunque a un’operazione di derivazione, le lingue endocentriche a una di composizione.

#### 4. Conclusioni: complementarità e isomorfia

Lo schema seguente rappresenta in modo succinto la complementarità nella creazione di lessemi non canonici (“inversi”), cioè la lessicalizzazione secondaria:

Tabella 2

	<b>Verbo (V)</b>	<b>Sostantivo (N)</b>
<b>Esocentrico</b>	composizione V + [MANIERA]	Derivazione N + [FUNZIONE]
<b>Endocentrico</b>	derivazione V + [DIREZIONE]	composizione N + [CONFIGURAZIONE]

Si tratta di tendenze maggiori, e come detto di modelli canonici – non di regole fisse, anche le lingue romane conoscono e utilizzano, ben inteso, la composizione nominale come visto qui sopra (per es. *portacenere*).

L'immagine completa della problematica trattata è esposta nella tabella 3:

Tabella 3

<b>Lessicalizzazione:</b>		<b>Primaria</b>		<b>Secondaria</b>
<b>Esocentrico</b>	<b>Verbo</b>	idea ( <i>uscire</i> )	→	immagine ( <i>uscire correndo</i> )
	<b>Sostantivo</b>	immagine ( <i>tavola</i> )	→	idea ( <i>scrivania</i> )
<b>Endocentrico</b>	<b>Verbo</b>	immagine ( <i>laufen</i> )	→	idea ( <i>hinauslaufen</i> )
	<b>Sostantivo</b>	idea ( <i>Kanne</i> )	→	immagine ( <i>Glaskaffeekanne</i> )

Ma di più, per ottenere un'astrazione nella lessicalizzazione secondaria, cioè l'espressione di un'idea, si utilizza la derivazione – che è un processo astratto. È il caso del sostantivo esocentrico, che per esprimere la idea di [FUNZIONE] aggiunge un suffisso (*sale* → *saliera*), e del verbo endocentrico, che per esprimere l'idea di [DIREZIONE] aggiunge un satellite (*laufen* → *hinaus-laufen*), cioè una derivazione. Per ottenere una concretizzazione, cioè l'espressione di un'immagine, si utilizza la composizione, processo concreto. Il verbo esocentrico, per esprimere l'immagine di [MANIERA], si combina con un altro verbo (*venire correndo*), e il sostantivo endocentrico, per esprimere l'immagine di [CONFIGURAZIONE], si combina con un altro sostantivo (*Glas-kaffeekanne*).

Al di là della complementarità fondamentale fra i due tipi, si può dunque constatare una isomorfia, o addirittura una iconicità, fra il senso prodotto e il procedimento grammaticale adottato:

Tabella 4

	<b>Senso</b>	<b>Procedimento grammaticale</b>
<b>Composizione</b>	concreto	concreto
<b>Derivazione</b>	astratto	astratto

Per ottenere un senso concreto, cioè l'espressione di un'immagine, si adotta, nella lessicalizzazione secondaria, un procedimento concreto, la composizione: il verbo esocentrico (tipo romanzo: *venire correndo*), il sostantivo endocentrico (tipo germanico: *Glaskaffeekanne*). Per ottenere un senso astratto nella lessicalizzazione secondaria, l'espressione di un'idea, si adotta un procedimento astratto, la derivazione: il sostantivo esocentrico (tipo romanzo: *saliera*), il verbo endocentrico (tipo germanico: *hinauslaufen*).

La composizione è detto processo concreto perché combina due o più elementi lessicali – senza peraltro specificare la relazione fra i lessemi così messi in rapporto, ma anche con tre componenti (ted. *Glas-kaffeekanne*), la composizione rimane purtroppo sempre binaria, cfr. Benveniste (1974), Baron (2000; 2003). La derivazione dall'altro lato è caratterizzata come processo astratto perché il satellite o l'affisso (pre- o suffisso) aggiunge un elemento grammaticale poco specifico come il suffisso italiano *-iera*, che denota semplicemente un rapporto di [FUNZIONE], per esempio *caffè, sale* → *caffettiera, saliera*.

Ecco la complementarità fondamentale fra i due tipi e all'interno di ciascuno di essi, la complementarità fra verbi e sostantivi – e l'isomorfia fra i procedimenti grammaticali messi in opera e il senso da produrre nella lessicalizzazione secondaria.

## Bibliografia

- Baron, I. 2000. La lexicalisation des noms composés déverbaux. Traits généraux de la composition nominale. In I. Korzen & C. Marellò (eds), *Argomenti per una linguistica della traduzione, Gli argomenti umani 4*. Alessandria: Edizioni dell'Orso, 43-52.
- Baron, I. 2003. Catégories lexicales et catégories de pensée : une approche typologique du danois et du français. In M. Herslund (ed.), *Aspects linguistiques de la traduction*. Bordeaux: Presses Universitaires de Bordeaux, 29-53.
- Baron, I. & Herslund, M. 2005. Langues endocentriques et langues exocentriques. Approche typologique du danois, du français et de l'anglais. In M. Herslund & I. Baron (eds), *Le génie de la langue française. Perspectives typologiques et contrastives. Langue française* 145: 35-53.
- Benveniste, E. 1974. Fondements syntaxiques de la composition nominale. In *Problèmes de linguistique générale II*. Paris : Gallimard, 145-162.
- Herslund, M. 2000. Tipologia grammaticale e tipologia lessicale. In I. Korzen & C. Marellò (eds), *Argomenti per una linguistica della traduzione, Gli argomenti umani 4*. Alessandria: Edizioni dell'Orso, 11-18.
- Herslund, M. 2003. Pour une typologie lexicale. In M. Herslund (ed.), *Aspects linguistiques de la traduction*. Bordeaux: Presses Universitaires de Bordeaux, 13-27.

- Herslund, M. 2007. Lexikalische Typologie – Dänisch und Französisch als endo- bzw. exozentrische Sprachen. In J. Rijkhoff (ed.), *Linguistic Typology. Tidsskrift for Sprogforskning* 5: 1-13.
- Herslund, M. 2009. Dal linguaggio alla cultura. Presentazione del progetto danese. In I. Korzen & C. Lavinio (eds), *Lingue, culture e testi istituzionali*. Firenze: Franco Cesati, 17-29.
- Jakobson, R. 1959. On linguistic aspects of translation. In R. A. Bowers (ed.), *On Translation*. Cambridge Mass.: Harvard University Press, 232-239.
- Korzen, I. 2005. Lingue endocentriche e lingue esocentriche: lessico, testo e pensiero. In I. Korzen & P. D'Achille (eds), *Tipologia linguistica e società*. Firenze: Franco Cesati, 31-54.
- Talmy, L. 1985. Lexicalization patterns: semantic structure in lexical forms. In T. Shopen (ed.), *Language Typology and Syntactic Description III*. Cambridge: Cambridge University Press, 57-149.

# KANTIAN GRAMMAR APPLIED TO FRENCH, DANISH AND ENGLISH

Hanne Korzen

Copenhagen Business School

## 1. Prologue

Many linguists refer to Kant, but they do not really seem to take him seriously. I will try to show that a little closer look at Kant's cognitive model might yield insight into certain important aspects of the syntactic-semantic constitution of the sentence in different languages. In the present work, I am especially concerned with the difference between space and time adjuncts on the one hand, and causal adjuncts on the other hand.

## 2. Systematic differences between French and Danish

Any Danish student of French is painfully aware of the enormous differences that separate the two languages. In order to specify, briefly, these differences, I will rely on the typological studies conducted by Herslund & Baron (e.g. Herslund 1997; Herslund & Baron 2003; Herslund & Baron 2005) and, as regards Italian, by Iørn Korzen (e.g. 1998; 2005; 2007).

Herslund & Baron (2003: 40) point out that “nouns and verbs are complementary in Danish and French with respect to the information density of the lexemes”. Thus, in the “endocentric”<sup>1</sup> language Danish – and in the other Germanic languages except English<sup>2</sup> – the information is concentrated in the verb, the centre of the clause (hence the term endocentric), whereas the information, in the “exocentric”<sup>3</sup> language French – and in the other Romance languages<sup>4</sup> – rather than

---

<sup>1</sup> Corresponding to Talmy's *Satellite-Framed Languages* (2000: 221).

<sup>2</sup> Due to the long French Influence, English is something between the Germanic and the Romance languages (cfr. Herslund & Baron 2005: 42ff). See below.

<sup>3</sup> Corresponding to Talmy's *Verb-Framed Languages*.

Cresti E., Korzen I. (eds), *Language, Cognition and Identity. Extensions of the endocentric / exocentric language typology* © 2010 Firenze University Press.

being localised in the clause-centre is distributed onto the nouns (hence the term *exocentric*, op. cit.: 31).

This difference can be illustrated by examples such as (1) and (2):

- |     |   |  |
|-----|---|--|
| (1) | L'oiseau <i>entre</i> dans la pièce.<br>[The bird enters the room.]<br>Le camion <i>entre</i> dans la cour.<br>[The lorry enters the courtyard]<br>Le bateau <i>entre</i> dans le port.<br>[The boat enters the port] | Fuglen <i>flyver</i> ind i værelset.<br>[The bird flies into the room.]<br>Lastvognen <i>kører</i> ind i gården.<br>[The lorry drives into the courtyard]<br>Båden <i>sejler</i> ind i havnen.<br>[The boat sails into the port] |
| (2) | Ræk mig lige <i>kanden</i> .<br><br>[Hand me the jug]   | Passez-moi la <i>cruche/le pichet/</i><br><i>le pot/le broc</i> , svp.<br>[Please, hand me the jug] <sup>5</sup> .   |

As (1) shows, it is often necessary to use a whole series of Danish movement verbs in order to translate one single French movement verb (depending on the context), the Danish verb, as opposed to the underspecified French verb, being obliged to express the *MANNER*<sup>6</sup> (cfr. Baron & Herslund 2003: 32). And as shown by (2), in many cases, one must use a whole series of French nouns in order to translate one single Danish noun (depending on the context), the French noun, as opposed to the underspecified Danish noun, being obliged to express the *CONFIGURATION*<sup>7</sup> (op. cit.: 38). In order to get the exact meaning, it is often necessary to qualify the very abstract Danish nouns and make them precise by a modifier in a morphologically composite structure (cfr. Herslund & Baron 2003: 36), whereas the most common nouns in French have the shape of monomorphemic lexemes with no need for further specification (loc. cit.). Herslund & Baron illustrate this by the following diagram:

---

<sup>4</sup> Cfr. especially I. Korzen (1998; 2005).

<sup>5</sup> *Larousse Dictionnaire Anglais* gives the following translations: *cruche*: jug UK, pitcher US; *pichet*: jug UK, pitcher US; *pot*: pot, jar, jug UK, pitcher US; *broc*: jug. So it seems that *jug* can be used in all the senses.

<sup>6</sup> Whereas the French verb lexicalises only the *PATH* component.

<sup>7</sup> Whereas the Danish noun only lexicalises the feature *FUNCTION*.

(3)	Danish	French
	stol	[ _ ]
	køkken-, spisestuestol	chaise
	[kitchen-, dining room- _]	
	Lænestol	fauteuil
	[lean- _]	
	korstol	stalle
	[choir- _]	
	prædikestol	chaire (church)
	[preach- _]	

(op. cit.)

Thus, whereas the Danish clause appears as a concrete relation between abstract entities, the French clause is an abstract relation between concrete entities (op. cit.: 31).

As mentioned above, English is not a typical Germanic language. Strongly influenced by French, this language shares properties with both the Germanic and the Romance languages. Thus in English, the genuine Germanic verbs behave roughly like the Danish verbs, whereas the – rather small – group of verbs that English borrowed from French, behave almost<sup>8</sup> like the Romance verbs (cfr. Herslund & Baron 2005: 42). This should appear clearly from the English translations of (1).

Nouns, too, sometimes behave like the Danish nouns (cfr. 4a) and sometimes like the French (cfr. 4b):

(4a)	Danish	English	French
	<i>hus</i>	<i>house</i>	[ _ ]
	(beboelses) <i>hus</i>	(dwelling) <i>house</i>	maison
	<i>lysthus</i>	<i>summerhouse</i>	pavillon
	<i>domhus</i>	<i>courthouse</i>	palais (de justice)
	(...)		

(4b)	Danish	English	French
	<i>sygehus</i>	hospital	hôpital
	<i>hundehus</i>	kennel	niche
	<i>kompashus</i>	binnacle	habitacle

(Herslund & Baron 2005: 50)

---

<sup>8</sup> Some of them have become transitive: *Entrer dans la pièce* → *enter the room* (cfr. Herslund & Baron 2005: 43).

This remarkable difference with respect to lexicalisation within the two major word classes seems to be – directly or indirectly – connected to important differences at other levels of the linguistic system such as: verbal morphology (cfr. for instance I. Korzen 2005: 42–44), syntax (cfr. for instance H. Korzen 2003; 2005; 2006), text structure (cfr. for example I. Korzen 2005; 2007; 2009; Lundquist 2005; 2009).

### 3. Some striking resemblances between French and Danish: the case of causal adjuncts

#### 3.1 Point of departure: the strange behaviour of ‘pourquoi’; the universality of the notion of cause

In some respects, however, the two languages are strikingly alike. In a series of publications (cfr. H. Korzen 1983; 1985; 1990), I have investigated the French interrogative word ‘pourquoi’, which, contrary to the other interrogative words, cannot be followed by stylistic inversion (but has to be combined with clitic inversion)<sup>9</sup>, and I have tried to explain why ‘pourquoi’ (‘why’), which functions as a causal adjunct, behaves differently from ‘quand’ (‘when’) and ‘où’ (‘where’), adjuncts of time and place, respectively. During this exploration, it struck me that most of the peculiarities attached to the causal adjuncts in French – and the way they differed from time and place adjuncts – were exactly the same in Danish, even though the inversion rules are quite different in the two languages.

In the mentioned works, I explained the difference between causal adjuncts and time and place adjuncts by postulating a different degree of attachment to the verb, as we shall see in 3.3.1 below. But why this difference in attachment? And why do causal adjuncts behave in the same way in two languages, which differ so much in other respects?

The notion of CAUSE is one of the “Semantic Primitives” laid down by Wierzbicka (1996: 70, 137, 186 ff). Referring to Kant, Wierzbicka remarks that

causation – with time and space – constitutes one of the basic categories of human cognition; it is not a category that we learn from experience but one of the categories which underlie our interpretation. (1996: 70)

---

<sup>9</sup> *Qui* est cette jolie fille? ‘Who is that beautiful girl?’ – *Que* fera Jean-Michel? ‘What will Jean-Michel do?’ – *A quoi* pensait Jean-Michel? ‘What was Jean-Michel thinking of?’ – *Quand* reviendra votre belle-mère? ‘When will your mother-in-law come back?’ – *Où* est garée votre voiture? ‘Where is your car parked?’ – *Comment* va votre fils? ‘How is your son?’ – *\*Pourquoi* pleure votre fils? → *Pourquoi* votre fils pleure-t-il? ‘Why is your son crying?’



and she points out that «data from language acquisition, as well as from cross-cultural semantics, are consistent with Kant's view» (ib.)<sup>10</sup>.

However, the fact that the three notions of time, space and cause are all universal semantic primitives, «irreducible categories of human language and cognition» (cfr. Wierzbicka 1996: 71), fails to explain why, then, the expressions for cause behave so differently from those expressing time and place. In order to explain this, we have to take a closer look at Kant's cognitive model.

### 3.2 Kant revisited

In his *Critique of Pure Reason* (Kant 1781)<sup>11</sup>, where he sets out to examine the foundations of human knowledge, Kant places the role of the human subject or “knower” at the center of our inquiry into nature, pointing out that all objects about which the mind can think must conform to its – rather limited – manner of thought<sup>12</sup>. As we can never escape the innate constraints of our minds, we must deal with them and accept that it is impossible to philosophise about things as they are, independently of us. Thus Kant makes a clear distinction between things as they appear to us as human beings, which are appearances in space and time, and the thing-in-itself (“das Ding an sich”), which we cannot ever come to know. Kant characterises the shift in point of view that made him focus on the human cognitive apparatus rather than the “outer world”, as his “Copernican Revolution”, because he attempted to reverse the mind-world relationship just as Copernicus had reversed the sun-earth relationship.

Kant points out in the *Introduction* that

there are two sources of human knowledge (which probably spring from a common, but to us unknown root), namely sensibility (“Anschauungsformen”)<sup>13</sup> and understanding (“Verstandesbegriffe”). By the former, objects are *given* to us, by the latter, they are *thought*. (*Critique*: 45)

Thus, in the first place, it is a matter of the aptitude to capture by the senses (sensibility), and in the second place the aptitude to interpret (understanding) what

---

<sup>10</sup> Wierzbicka remarks, on the one hand, that apparently all languages have a lexical exponent of causation (whether it is a conjunction like ‘because’, a noun like ‘cause’, or an ablative suffix), and on the other hand, that ‘because’-sentences appear quite early in children’s speech «despite the highly abstract and “non-empirical” character of the concept of causality» (op. cit.).

<sup>11</sup> From now on “*Critique*”. The side numbers refer to the English edition from 1993.

<sup>12</sup> Both rationalism and empiricism had overlooked the fact that the human mind is limited, and that it can experience and imagine only within certain constraints.

<sup>13</sup> The English word ‘sensibility’ does not really render the sense of ‘Anschauung’, but it is difficult to find another word.

we have sensed e.g. as a relation of cause and effect. Kant considers both aptitudes as necessary (and inherent in man):

Neither of these faculties has a preference over the other. Without the sensible faculty no object would be given to us, and without the understanding no object would be thought. Thoughts without content are empty; intuitions without concepts, blind. Hence it is as necessary for the mind to make its concepts sensible (that is, to join to them the object in intuition), as it is to make its intuitions intelligible (that is, to bring them under concepts). Neither of these faculties can exchange its proper function. Understanding cannot intuit, and the sensible faculty cannot think. In no other way than from the united operation of both, can knowledge arise (*Critique*: 69–70)

But Kant explicitly warns us against mixing them up:

But no one ought, on this account, to overlook the difference of the elements contributed by each; we have rather great reason carefully to separate and distinguish them. We therefore distinguish the science of the laws of sensibility, that is, Aesthetic, from the science of the laws of the understanding, that is, Logic (*Critique*: 69–70)

Besides, the distinction between the two faculties appears clearly from the organisation of the book. They are both discussed in the first (and longest) part: “Transcendental Doctrine of Elements”. This part, which deals with the fundamental building blocks of experience, is divided into two chapters: 1) “Transcendental Aesthetic” and 2) “Transcendental logic”. Let us take a brief look at these chapters.

“Transcendental Aesthetic” deals with the two forms of sensibility that are a priori conditions for any possible experience, namely Space and Time. Space is a necessary presupposition for being able to observe at all. Or, as Kant puts it himself:

Space [...] is a necessary representation a priori, which serves for the foundation of all external intuitions. (*Critique*: 50)

It is a universally valid and necessary (i.e., a priori) truth that everything must necessarily be found at some place or other. (cfr. Hartnack 1967: 18). As for Time, Kant points out that

Time is nothing else than the form of the internal sense, that is, of the intuitions of ourselves and of our internal state. (*Critique*: 56)

We cannot experience anything without presupposing time. One cannot imagine a world that is not in time, i.e., a world where nothing happens either before, at the

same time as, or after something else. It is a universally valid and necessarily true proposition that every event and process occurs at a given moment in time and that every process takes a certain time (cfr. Hartnack 1967: 23).

“Transcendental logic” treats the fundamental concepts<sup>14</sup> of understanding, which Kant calls categories<sup>15</sup>. The categories synthesise the random data of the sensory manifold into intelligible objects. There are twelve categories, among which we find the relation of “Causality and Dependence (cause and effect)” (*Critique*: 85)<sup>16</sup>:

- |     |   |   |
|-----|---|---|
| (5) | 1.<br><i>Quantity</i><br>unity<br>plurality<br>totality   |   |
|     | 2.<br><i>Quality</i><br>reality<br>negation<br>limitation   | 3.<br><i>Relation</i><br>substance<br><u>causality</u><br>community |
|     | 4.<br><i>Modality</i><br>possibility–impossibility<br>existence–nonexistence<br>necessity–contingency |   |

(Hartnack 1967: 45; *Critique*: 85)

The categories under ‘quantity’ and ‘quality’ Kant calls “the mathematical categories”; these categories indicate the conditions for making judgments about objects in space and time. The categories under ‘relation’ and ‘modality’ Kant calls “the dynamic categories”; these categories indicate how an object is determined in relation to other objects (ibid.)

As it should appear from what has been said above, Cause, a dynamic category, differs completely from Space and Time, the two forms of sensibility. And one

---

<sup>14</sup> These concepts are a priori concepts, i.e., concepts that are not formed by abstracting from experience (cfr. Hartnack 1967: 32).

<sup>15</sup> These could be discovered, according to Kant, by an examination of the forms of the synthetic a priori judgments.

<sup>16</sup> Cfr. Yakira (1994: 114-15): «La causalité est une des douze catégories qui déterminent les formes principales par lesquelles le divers donné par les sens est unifié, et par là-même, devient déterminé et objectif ou, plus simplement, reçoit un sens».

could say that it is, in a certain sense, less fundamental than these. Moreover, already before setting out to explain the “Transcendental Doctrine of Elements”, Kant explicitly states:

So far as the faculty of sensibility may contain representations a priori, which form the conditions under which objects are given, in so far it belongs to transcendental philosophy. The transcendental doctrine of sensibility must form the *first* part of our science of elements, because *the conditions under which alone the objects of human knowledge are given, must precede those under which they are thought.* (*Critique*: 45)

Thus, Kant does not place ‘Cause’ at the same level as ‘Space’ and ‘Time’, whatever different linguists seem to have meant. Directly connected to the senses, ‘Space’ and ‘Time’ constitute the preconditions for any possible experience and function as “the scene” where experiences appear (cfr. Thomsen 1964: 270). ‘Cause’ adds, as it were, an explanation to the sensed phenomena.

In the following section, we shall see that this difference between Space and Time on the one hand, and Cause on the other hand have quite radical consequences for the syntactic-semantic constitution of the sentence.

### 3.3 The cause adjunct and the other sentence members

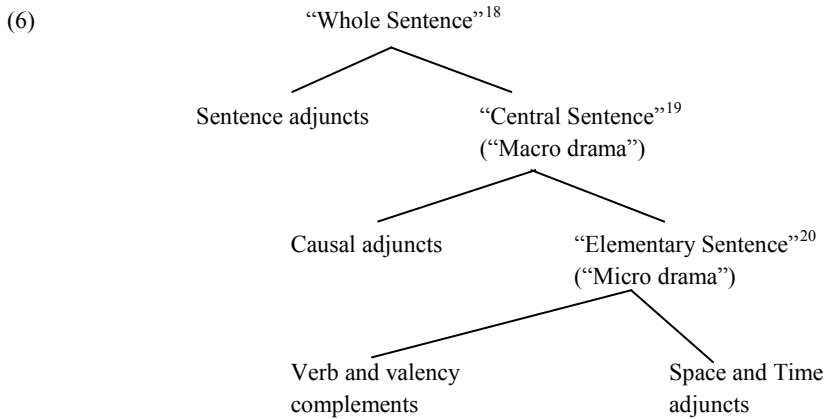
In H. Korzen (1983; 1985; 1990), I examined the special behaviour of cause adjuncts compared to the other sentence members, especially space and time adjuncts, from which it distinguishes itself in several respects. Below I will show a small representative sample of the characteristic properties of these syntactic functions.

#### 3.3.1 *The hierarchical model*

The point of departure of my description is the following hierarchical model<sup>17</sup>, where the causal adjunct occupies an intermediary level between the sentence adjuncts (e.g. *heureusement*) and constituents capable of triggering stylistic inversion (cfr. (6) below):

---

<sup>17</sup> (6) is not a syntactic tree but a graphic representation of the hierarchical organisation of the elements that one can find in a sentence, and the terms “Whole Sentence”, “Central Sentence” and “Elementary Sentence” mean: “elements likely to be found in the Whole Sentence etc.”.



The Central Sentence denotes the proper content, i.e. the part that is asserted, as opposed to the sentence adjuncts, which are merely “shown” (in the sense of Wittgenstein, cf. Nølle 1999). In H. Korzen (1985), I said that the sentence adjuncts were “periphery elements”, and I will use that term below. The Elementary Sentence, which contains the verb and its valency complements and possibly space and time adjuncts, denotes the situation which is the centre of interest. In order to abbreviate, I will use the term “elementary constituents” in order to talk of the valency complements and the space and time adjuncts. These sentence members share several properties by which they distinguish themselves from the causal adjuncts.

### 3.3.2 Differences between the causal adjuncts and the elementary constituents

#### – The “essential” character of the elementary constituents: quantifiability

The elementary constituents can be considered “essential” because they are necessary in order that one can say that an event has taken place. If you negate the existence of one of them by means of an expression signifying “zero”, it amounts to negating the whole proposition (cf. Cornulier 1974:161). This is what we see in (7a-e):

(7a) *Personne* ne chante. (Subject)

[Nobody sings]

(7b) *Jeanne* ne mange *rien*. (Object)

[Jane eats nothing]

<sup>18</sup> In Korzen (1983; 1985; 1990): “Phrase entière”.

<sup>19</sup> In Korzen (1983; 1985; 1990): “Phrase centrale”.

<sup>20</sup> In Korzen (1983; 1985; 1990): “Phrase élémentaire”.

- (7c) Je ne donnerai ce livre *à personne*. (Indirect object)  
[I will not give this book to anyone]
- (7d) Michèle ne travaille *nulle part*. (Space adjunct)  
[Michelle does not work anywhere]
- (7e) Michèle ne travaille *jamais*. (Time adjunct)  
[Michelle never works]

which signifies: (7a) ‘There is no singing at all’, (7b) ‘Jane does not eat at all’, (7c) ‘I will not make a present of this book at all’, (7d) and (7e) ‘Michelle does not work at all’. In all these constructions the negated constituents are negation words which form the second part of the negation.

The causal adjunct does not behave like that. Negating the cause does not amount to negating the whole proposition:

- (8) Michèle pleure sans raison.  
[Michelle cries without reason]

(8) does not mean: ‘Michelle does not cry at all’. In fact, it happens very often that someone talks of a phenomenon while he maintains that no other phenomenon provoked it<sup>21</sup>. It is significant that there is no negative word in French that denotes ‘for no reason’ corresponding to *jamais* (‘never’) and *nulle part* (‘nowhere’). In Danish and English too, there are special negative words (the Danish negative words being derived from the Old Norse *engi* ‘nothing’, and the English equivalents being a combination of the negation and a noun or an adverb) corresponding to the French negative words:

- (9) ingen = nobody; intet = nothing; intetsteds = nowhere; ingensinde/aldrig = never

But there is no corresponding expression denoting ‘for no reason’.

It might seem difficult to reconcile this with Kant’s theory, according to which «causation – with time and space – constitutes one of the basic categories of human cognition» (Wierzbicka 1996: 70, cfr. 3.1 above). But here, one must bear in mind that Kant did not put the three notions at the same level. Space and Time are more fundamental, as they are directly connected to the senses. What I really mean when I say ‘without any reason’ is obviously ‘without any *apparent* reason’, i.e. ‘for a reason unknown to us / a reason to which we have no access’. The idea of cause does not arise until we are presented (preferably several times) with *two or more* events following one after the other. Then the human mind will tend to conclude that

---

<sup>21</sup> More than 2 000 000 hits on Google for *sans [aucune] raison*, and more than 52 000 000 for *without [any] reason*, 19/9 2009.

there is a causal relation between them<sup>22</sup>. If you “remove” the cause, the world becomes absurd or inexplicable. But if you “remove” space and time, you pull the rug from under your feet, and there will be nothing left.

We have just seen that it is possible to negate the whole proposition by negating the existence of one of the elementary constituents by means of an expression signifying “zero”. On the other hand, it is possible to “reinforce” it by “reinforcing” one of the elementary constituents. Thus, one could say that the examples in (11a-e) express a larger quantity of action than the examples in (10a-e):

- (10a) *Jean-Michel* chante.  
[Jean-Michel is singing]
- (10b) Jean-Michel a embrassé *Maryse*.  
[Jean-Michel kissed Maryse]
- (10c) Jean-Michel a donné des cadeaux à *Sophie*.  
[Jean-Michel gave a present to Sophie]
- (10d) On a manifesté à *Paris*.  
[They demonstrated in Paris]
- (10e) Michel vient ici *le mardi*.  
[Michel comes here every Tuesday]
- (11a) *Jean-Michel et Pierre* chantent. (‘il y a deux fois plus de chant’)  
[Jean-Michel and Pierre are singing. (‘there is twice as much singing’)]
- (11b) Jean-Michel a embrassé *Maryse et Mathilde*. (‘il y a deux fois plus de baisers’)  
[Jean-Michel kissed Maryse and Mathilde. (‘there is twice as many kisses’)]
- (11c) Jean-Michel a donné des cadeaux à *Sophie et à Irène*.  
[Jean-Michel gave gifts to Sophie and Irène]
- (11d) On a manifesté à *Paris et à Lyon*. (‘on a manifesté à deux endroits’)  
[They demonstrated in Paris and Lyon’ (‘they demonstrated in two places’)]
- (11e) Michel vient ici *le mardi et le jeudi*. (‘Michel vient ici deux fois par semaine’)  
[Michel comes here every Tuesday and every Thursday.  
(‘Michel comes twice a week’)]

As for the causal adjunct, increasing the number of causes does not result in a “bigger amount of action”, which remains the same in (13) as in (12):

- (12) Michel a embrassé Maryse *parce qu’il l’aime*.  
[Michel kissed Maryse because he loves her]
- (13) Michel a embrassé Maryse *parce qu’il l’aime et que le soleil brille*.  
[Michel kissed Maryse because he loves her and the sun is shining]

---

<sup>22</sup> The notion of Cause is not empiric (Hume).

Sometimes a whole lot of reasons are given for the same situation:

- (14) Je ne suis pas allé en classe parce qu'il fait froid, parce qu'il pleut, parce qu'il neige, parce qu'il gèle, parce qu'il y a du brouillard, parce que le ciel est gris, parce qu'il vente et qu'il grêle (Ionesco: *Le bon et le mauvais temps*: 130)<sup>23</sup>  
 [I did not go to school because it is cold, because it is raining, because it is snowing, because it is freezing, because it is foggy, because the sky is gray, because the wind is blowing and because it is hailing]

I will say that the elementary constituents are “quantifiable”. This quantifiability has important morpho-syntactic consequences as we shall see below.

– *The type ‘n’importe qui’*

It is possible to “reinforce”, as it were, the truth value of a proposition by saying that the proposition is true whenever or wherever it takes place, and whatever be the entities that participate in it. In order to express this phenomenon, there are pronominal phrases corresponding to the quantifiable constituents like:

- (15) Fr. *n’importe qui*, *n’importe quoi*, *n’importe où*, *n’importe quand*  
 Eng. *whoever*, *whatever*, *wherever*, *whenever*  
 Da. *hvem* som helst, *hvad* som helst, *hvor* som helst, *når* som helst

Whereas *pourquoi* and its Danish and English equivalents cannot be part of this kind of expression:

- (16) \**n’importe pourquoi*, \**hvorfor* som helst, \**whyever*<sup>24</sup>

– *Clause combining*

Two situations can be connected by the fact that they share the same time, the same space, the same object, the same person, etc. Thus, they share one of the “essential” aspects. In order to express this, there are combinations of subordinate clauses and their main clauses like (17a-d) in French, (18a-d) in English and (19a-d) in Danish<sup>25</sup>:

<sup>23</sup> Thanks to Birgitte Regnar for supplying me with this wonderful example.

<sup>24</sup> It is true that you can have expressions like: *pour n’importe quelle raison*, *for whatever reason*, *af hvilken som helst grund*. But it is significant that it is not the ordinary, unmarked wh-word that is used in the expression.

<sup>25</sup> In the Danish tradition we call the shared element “common member”, cfr. Diderichsen (1946/1966).



- (17a) Tu peux venir *quand* tu voudras.  
 (17b) Tu peux dîner *où* tu voudras.  
 (17c) Tu peux faire *ce que* tu voudras.  
 (17d) Tu peux épouser *qui* tu voudras.
- (18a) You can come *when(ever)* you want.  
 (18b) You can dine *where(ver)* you want.  
 (18c) You can do *what(ever)* you want.  
 (18d) You can marry *who(ever)* you want.

- (19a) Du kan komme *når* du (*end*) vil.  
 (19b) Du kan spise *hvor* du (*end*) vil.  
 (19c) Du kan gøre *hvad* du (*end*) vil.  
 (19d) Du kan gifte dig med *hvem* du (*end*) vil.

In (17)-(19) the time, place, etc. are non-specific. But the same “share” is possible when they are specific:

- (20) Pierre est parti *quand* Charles est venu.  
 [Pierre left when Charles arrived]
- (21) Michèle habite *dans la maison où* Jacques travaille.  
 [Michel lives in the house where Jacques works]

No construction corresponding to (17)-(21) denotes that two events happen for the same reason, so the causal conjunction is excluded from (22), corresponding to (17)-(19):

- (22) Fr. \*Tu peux venir *pourquoi* tu voudras.  
 Eng. \*You can come *why(ever)* you want.  
 Da. \*Du kan komme *hvorfor* du vil.

And a construction like (23), which contains a causal clause, does not express that two events happen for the same reason, but that the one denoted by the causal clause explains the other one:

- (23) Pierre est parti *parce que* Charles est venu.  
 [Peter left because Charles came]

Of course, it happens that two (or more) situations can be explained by one and the same reason. Say Paul stayed at home because it was snowing (he hated snow), but Jacques went out because it was snowing (he loved snow). Theoretically, it should be possible to imagine a construction like (24), but it just “sounds” crazy:

- (24) ?Paul est resté chez lui pour la même raison que celle pour laquelle Jacques est sorti.  
[Paul stayed at home for the same reason as the reason for which Jacques went out]

No language seems to use such a construction. What is “shared” in constructions like (17)-(21) is always one of the essential aspects, which are, explicitly or implicitly, “present”<sup>26</sup> in all situations. They are part of any situation. These possibilities of clause combining are probably the most spectacular consequence of the difference between the time and space adjuncts on the one hand, and the causal adjunct on the other.

– *Periphery properties*

Being outside the Elementary Sentence, the causal adjuncts share certain properties with the sentence adjuncts. They can for instance be outside the scope of negation and combine directly with the negative adverb corresponding to *not*:

- (25) Fr. *Pourquoi* pas ?  
Eng. *Why* not?  
Da. *Hvorfor* ikke ?

The causal adverb, in (25) occupies exactly the same place as *unfortunately* in (26):

- (26) Fr. *Malheureusement* pas !  
Eng. *Unfortunately* not!  
Da. *Desværre* ikke !

That position is excluded for space and time adjuncts:

- (27) Fr. *\*Quand* pas ? *\*Où* pas ?  
Eng. *\*When* not? *Where* not?  
Da. *\*Hvornår* ikke ? *\*Hvor* ikke ?

The causal adjuncts may also – just like the sentence adjuncts but contrary to space and time adjuncts – modify an elliptic sentence:

- (28) Fr. Michel va venir. – *Pourquoi* Michel? *\*Quand* Michel? *\*Où* Michel?  
Eng. Michel will come. – *Why* Michel? *\*When* Michel? *\*Where* Michel?  
Da. Michel kommer. – *Hvorfor* Michel? *\*Hvornår* Michel? *\*Hvor* Michel?

---

<sup>26</sup> They do not have to be physically there, but they are always implied.

The constructions in (28) are the same kind of constructions as (29), where the sentence adjunct *unfortunately* and its equivalents modify the elliptic construction:

- (29) Fr. Qui va venir? – Michel, *malheureusement*.  
 Eng. Who will come? – Michel, *unfortunately*.  
 Da. Hvem kommer? – Michel, *desværre*.

This difference between space and time adjuncts and the adjuncts which are outside the Elementary Sentence is probably due to the fact that an elliptic sentence already, implicitly, contains the other Elementary constituents (i.e. time and place adjuncts, etc.), for which reason it can only be modified by “external” sentence members.

### 3.3.3 *The causal adjunct and the elementary constituents: temporary conclusion*

It should appear from 3.3.2 that space and time adjuncts share important properties with the other Elementary constituents. All these constituents denote essential aspects of the situation which is the centre of interest, and because of that, they differ from all other constituents that can be found in a sentence. The causal adjunct, on the other hand, is excluded from this exclusive circle, and it shares several properties with the sentence adjuncts.

## 4. Conclusion

We must conclude that, as far as the properties examined in 3.3 are concerned, French, English and Danish have almost the same syntax. In all three languages, the space and time adjuncts differ fundamentally – and in the same ways – from the causal adjuncts, although these languages, being typologically different, differ from each other in many other important respects. This fits so nicely in with Kant’s cognitive model. The next step would be an investigation of some non indo-european languages<sup>27</sup>. But I think it is very probable that, due to the human cognitive apparatus, the discussed differences between the time and space expressions on the one side, and causal expressions on the other are universal.

Let us have a final look at the hierarchical model (6) in order to see in which respects the investigated languages differ, and in which respects they are alike. It seems that the differences concern above all the verb and its complements (i.e. on a rather “concrete level”). As Herslund & Baron and I. Korzen have shown, the different information density in the two major word classes, nouns and verbs,

---

<sup>27</sup> As a matter of fact, I already started this investigation, and it looks like Japanese and Hungarian behave roughly like the languages studied her. This will be the subject of my next publication.

seems to be directly or indirectly connected to important differences at other levels of the linguistic system. The resemblances are to be found at a more abstract level: expressions for Time, Space and Cause. Cause (together with some other “operators”, above all negation) are probably good candidates for the exclusive group of linguistic universals.

## Bibliography

- Baron, I. & Herslund, M. 2005. Langues endocentriques et langues exocentriques. Approche typologique du danois, du français et de l’anglais. In M. Herslund & I. Baron (eds), *Le génie de la langue française. Perspectives typologiques et contrastives. Langue Française* 145: 35-53.
- Cornu, J.-M. 2007. Emmanuel Kant (1724-1804).  
<http://216.239.59.104/search?q=cache:Fw4HvueIocclJ:www.cornu.eu.org/news/em>.
- Cornulier, B. de. 1974. *Pourquoi* et l’inversion du sujet non clitiques. In C. Rohrer & N. Ruwet (eds.), *Linguistische Arbeiten, Rohrer, Ruwet. – Actes du colloque Franco-allemand de Grammaire Transformationelle. I. Etudes de syntaxe*. Tübingen: Niemeyer, 139-63.
- Diderichsen, P. 1946<sup>1</sup>, 1966<sup>2</sup>. *Elementær Dansk Grammatik*. Copenhagen: Gyldendal.
- Hartnack, J. 1967. *Kant’s Theory of Knowledge*. Indianapolis/Cambridge: Hackett.
- Herslund, M. (ed.) 1997. *Det Franske Sprog. Kapitel I. Grundlag*. Copenhagen Business School.
- Herslund, M. & I. Baron. 2003. Language as World View. Endocentric and exocentric representations of reality. In I. Baron (ed.), *Language and Culture. Copenhagen Studies in Language* 29: 29-42.
- Herslund, M. & Baron, I. 2005. Présentation : typologie lexicale et au-delà. In M. Herslund & I. Baron (eds), *Le génie de la langue française. Perspectives typologiques et contrastives. Langue Française* 145: 3-9.
- Kant, E. 1781<sup>1</sup>, 1787<sup>2</sup>. *Kritik der reinen Vernunft*. English translation based on Meikeljohn, edited by Vasilis Politis. Everyman 1993.
- Korzen, H. 1983. Réflexions sur l’inversion dans les propositions interrogatives en français. In M. Herslund, O. Mørdrup & F. Sørensen (eds), *Analyses grammaticales du français. Etudes publiées à l’occasion du 50<sup>e</sup> anniversaire de Carl Vikner. Revue Romane numéro spécial* 24: 50-85.
- Korzen, H. 1985. *Pourquoi* et l’inversion finale en français : étude sur le statut de l’adverbial de cause et l’anatomie de la construction tripartite. *Revue Romane numéro supplémentaire* 30.
- Korzen, H. 1990. Pourquoi *pourquoi* est-il différent? L’adverbial de cause et la classification des adverbiaux en général. In H. Nølke (ed.), *Classification des adverbes. Langue Française* 88: 60-79.

- Korzen, H. 2003. Attribut de l'objet et valence dérivée. Étude contrastive dano-française. In M. Herslund (ed.), *Aspect linguistiques de la traduction*. Bordeaux : Presses Universitaires de Bordeaux, 85-102.
- Korzen, H. 2005. Attributs directs et indirects en français, en danois et en anglais: différences typologiques et problèmes de traduction. In M. Herslund & I. Baron (eds), *Le génie de la langue française. Perspectives typologiques et contrastives*. *Langue Française* 145: 55-73.
- Korzen, H. 2006. Appositions and related constructions in French and Danish. In P. Colliander (ed.), *Appositions in Selected European Languages*. *Copenhagen Studies in Language* 33: 53-92.
- Korzen, I. 1998. *At tale om ting. Om substantiver og substantivgrupper betydning og brug – med særligt henblik på italiensk og dansk*. *Copenhagen Working Papers in LSP*. Copenhagen Business School.
- Korzen, I. 2005. Lingue endocentriche e lingue esocentriche: lessico, testo e pensiero. In I. Korzen & P. D'Achille, Paolo (a cura di), *Tipologia linguistica e società. Considerazioni inter- e intralinguistiche. Due giornate italo-danesi di studi linguistici*. Firenze: Franco Cesati, 31-54.
- Korzen, I. 2007. Mr. Bean e la linguistica testuale comparativa. Considerazioni tipologico-comparative sulle lingue romanze e germaniche. In M. Barbera, E. Corino & C. Onesti (a cura di), *Corpora e linguistica in rete*. Perugia: Guerra, 209-224.
- Korzen, I. 2009. Struttura testuale e anaphora evolutiva: tipologia romanza e tipologia germanica. In I. Korzen & C. Lavinio (a cura di), *Lingue, culture e testi istituzionali*. Franco Cesati. Firenze, 33-60.
- Lundquist, L. 2005. Noms, verbes et anaphores (in)fidèles. Pourquoi les Danois sont plus fidèles que les Français. In M. Herslund & I. Baron (eds), *Le génie de la langue française. Perspectives typologiques et contrastives*. *Langue Française* 145: 73-93.
- Lundquist, L. 2009. Adverbiaux initiaux en danois et en français: langue, texte, mentalité. In I. Korzen & C. Lavinio (a cura di), *Lingue, culture e testi istituzionali*. Firenze: Franco Cesati, 141-63.
- Nølke, H. 1999 (ed.). *Det Franske Sprog. Kapitel V. Modification 2*. Copenhagen Business School.
- Puech, M. 1990. *Kant et la causalité*. Paris: Librairie Philosophique J. Vrin.
- Thomsen, H. 1964. *Hvem tænkte hvad*. Copenhagen: Politikens Forlag.
- Vikner, C. 1978. Les auxiliaires négatifs : fonction et position. *Revue Romane* XIII, 1: 88-109.
- Wierzbicka, A. 1996. *Semantics. Primes and Universals*. Oxford: Oxford University Press.
- Yakira, E. 1994. *La causalité de Galilée à Kant*. Paris: Presses Universitaires de France.



# I VERBI GENERALI NEI CORPORA DI PARLATO. UN PROGETTO DI ANNOTAZIONE SEMANTICA CROSS-LINGUISTICA

Massimo Moneglia, Alessandro Panunzi  
LABLITA, Università di Firenze

## 1. Verbi generali di azione

1.1 Un'azione è un modello di modificazione del mondo da parte di un attore che può essere applicato a un insieme aperto di oggetti. Parallelamente ogni oggetto può sottostare a un insieme aperto di azioni. Ciò significa che le azioni sono entità ontologiche produttive. Più precisamente, le azioni possono essere considerate concetti proprio in quanto sono entità produttive rispetto ai loro oggetti. Ogni azione, diversamente, sarebbe un *hapax* (Bruner 1983; Minsky 1968; Piaget 1954; Tomasello 2003; Von Wright 1963).

La produttività dell'azione è riflessa a livello logico nel linguaggio dai verbi, che, allo stesso modo, possono applicarsi a insiemi aperti di argomenti, argomenti che sono a loro volta oggetto di molti verbi. Ma i predicati naturali sono più creativi della predicazione "logica". Per esempio se si considerano le istruzioni relative al predicato 'aprire', possiamo notare (vedi figura 1) che questo può essere applicato produttivamente a serie qualitativamente distinte di azioni.

Anche se in tutti i casi è stato utilizzato lo stesso verbo, l'azione non è la stessa. Per esempio l'azione di 'aprire una scatola' permette l'accesso a un contenuto interno, mentre nell'azione di 'aprire una finestra' dà accesso allo spazio esterno. Quando si apre un ombrello le relazioni spaziali non caratterizzano in nessun modo l'evento. Anche negli altri casi le proprietà focali dell'azione cambiano in maniera radicale, tanto che è possibile pensare a molte azioni diverse individuate dal verbo piuttosto che ad una sola.

Questo giudizio è confermato dalla produttività tipologica che le variazioni precedenti mostrano. Così come siamo capaci di giudicare che le variazioni in figura 1 producono azioni diverse, siamo anche capaci di dire che, al contrario, ci sono variazioni che *non* producono azioni qualitativamente diverse, ovvero che variando

l'oggetto non si produce una variazione di tipo di azione. In questo caso il tipo risulta produttivo al suo interno.



Figura 1. Variazione di 'aprire'

Per esempio, non si ha variazione di tipo di azione nelle serie seguenti, ovvero tra le istanze di ogni serie non c'è la diversità tipologica che si verifica tra le istanze di serie diverse:

- (1) Qualcuno apre una finestra / una porta / una tenda
- (2) Qualcuno apre la scatola / il cassetto / la cartella
- (3) Qualcuno apre il libro / il laptop / la carta geografica
- (4) Qualcuno apre l'ombrello, una biro, la serratura, la sedia a sdraio.
- (5) Qualcuno apre gli occhi / la bocca / le braccia / la sua mano /
- (6) Qualcuno apre una noce / un uovo / un pacchetto / una fialetta

Ciascun tipo è produttivo al suo interno ed è quindi da considerare, coerentemente alla proprietà della produttività che caratterizza il concetto di azione, un'azione indipendente a livello concettuale.

Dunque per un verbo come 'aprire' la relazione tra *azione*, in quanto entità ontologica, e *verbo*, in quanto entità linguistica, non è una relazione uno a uno. In altre parole, il linguaggio ordinario non rispecchia l'ontologia dell'azione. Chiamiamo *generale* un verbo che ha tale proprietà e *Variazione primaria* (o *verticale*) del verbo la serie dei tipi di atti che possono rientrare nella sua estensione. Chiamiamo *Variazione orizzontale* la variazione che si verifica all'interno di ciascun tipo e che ne costituisce la produttività.

Come vedremo più avanti, le azioni che sono più frequentemente oggetto di riferimento nell'uso linguistico sono categorizzate attraverso verbi generali. Ciò



produce una notevole indeterminatezza nel riferimento all'azione nei linguaggi naturali e rende problematici tutti i *task* che richiedono la sua disambiguazione. Ma la variazione dei verbi generali ammissibile nel lessico delle lingue non è conosciuta con precisione. La soluzione di questo problema è resa più urgente dal fatto che le variazioni dei verbi generali in lingue diverse non sono parallele, per cui si pone un serio problema per la traduzione (in particolare automatica) specificamente dovuto all'interpretazione di tali verbi. In questo lavoro illustreremo brevemente le linee teoriche su cui si svilupperà un progetto di ricerca e sviluppo dedicato alla formazione di una base di dati che contiene l'informazione necessaria alla disambiguazione dei verbi generali<sup>1</sup>.

1.2 La variazione di forma negli eventi denotati dai predicati naturali è un dato di fatto già compiutamente notato da Wittgenstein (1953). Wittgenstein sosteneva che la applicazione dei predicati naturali non è strettamente regolata, ma dipende da un lato dall'uso e dall'altro dalla libertà dei soggetti di applicare predicati adatti a designare un certo evento tipico anche a eventi nuovi ad esso somiglianti. L'estensione di un predicato si caratterizza in questo approccio per essere composta da famiglie di eventi tra loro simili, che corrispondono, grosso modo, ai tipi che abbiamo individuato e che secondo la tradizione si distinguerebbero in usi più centrali e più periferici del predicato stesso (Givon 1986).

Ciò corrisponde però solo in parte a quanto osservato in precedenza. È interessante rimarcare che l'effettivo rapporto tra centro e periferia dei concetti (Rosh 1978), quando è riscontrato nella variazione dei verbi generali, convalida l'idea che questi verbi hanno una variazione tipologica. E ogni tipo è in effetti una azione indipendente. Possiamo notare infatti che nessuno dei tipi di *aprire* è più prototipico dell'altro, e tutti possono essere riconosciuti come campi di applicazione centrali del verbo. In altri termini, non possiamo dire che uno dei tipi nella variazione è più 'aprire' dell'altro, e tutti potrebbero essere indifferentemente portati ad esempio di cosa intendiamo per 'aprire'. Sono tutti possibili prototipi.

Al contrario, all'interno della variazione orizzontale di ciascun tipo è in effetti possibile distinguere tra centro e periferia. Per esempio si può concordare con il fatto che 'aprire la finestra' è più centrale e prototipico di 'aprire la tenda', che 'aprire l'ombrello' è più centrale di 'aprire la sedia a sdraio', che 'aprire gli occhi' è più centrale di 'aprire la mano', eccetera. Dunque, se l'aver centro e periferia è una proprietà dei concetti naturali, allora i tipi di atti nella variazione di un verbo generale sono da considerare a tutti gli effetti concetti naturali, in quanto hanno produttività e la conseguente variazione disegna un centro e un periferia.

---

<sup>1</sup> Progetto IMAGACT, "Ontologia per immagini dell'azione per l'acquisizione di L2 in ambiente multilingue", finanziato dalla Regione Toscana sui fondi PAR FAS 2007-2013 Linea di azione 1.1.3.

## 2. Variazione primaria e marcata

Non è certo una novità notare che il significato dei verbi è estremamente variabile. La variabilità semantica del lessico, in particolare verbale, specificamente in relazione al contesto lessicale in cui questo compare, è stata chiara fin dagli esordi della lessicografia (Vocabolario degli Accademici della Crusca 1612)<sup>2</sup>. È necessario però operare almeno una distinzione in tale variabilità perché il suo studio consenta di acquisire dati relativamente alla competenza semantica e quindi consenta previsioni propriamente sulla produttività nell'applicazione dei predicati verbali in considerazione.

Osservare cosa riporta della variazione del verbo che abbiamo appena studiato (*to open*) una moderna ontologia (WordNet) consente di evidenziare semplicemente questo punto.

Tabella 1. *To open* in WordNet

N.	Synset	Definition	Example
1	open, open up	cause to open or to become open	<i>Mary opened the door</i>
2	open, open up	start to operate or function or cause to start operating or functioning	<i>open a business</i>
3	open, open up	become open	<i>The door opened</i>
4	open	begin or set in action, of meetings, speeches, recitals, etc.	<i>He opened the meeting with a long speech</i>
5	unfold, spread, spread out, open	spread out or open from a closed or folded state	<i>open the map</i>
6	open, open up	make available	<i>This opens up new possibilities</i>
7	open, open up	become available	<i>an opportunity opened up</i>
8	open	have an opening or passage or outlet	<i>The bedrooms open into the hall</i>
9	open	make the opening move	<i>Kasparov opened with a standard opening</i>
10	afford, open, give	afford access to	<i>the door opens to the patio</i>
11	open	display the contents of a file or start an application as on a computer	

Come si può notare, alcuni dei tipi precedentemente individuati ('aprire l'ombrello', 'aprire gli occhi') non sono presenti, mentre sono listate variazioni del verbo

<sup>2</sup> Vocabolario degli Accademici della Crusca (1612): «Pigliare: [...] diversifica questo verbo i significati, secondo le parole, che l' accompagnano».

(*synsets* nel lessico di WordNet) nelle quali, come fanno anche i dizionari, si riportano una varietà di usi che, al contrario non sono stati presi in considerazione nel nostro studio della variazione di “aprire”. In particolare, usi astratti (4, 6, 7, 9) e usi specifici locativi (10).

Ciò non è un caso. La variazione individuata in WordNet non è infatti riferita specificamente al concetto di *tipo azionale* e riporta insieme sia parte della tipologia di azioni fisiche estese dal verbo, sia usi fraseologici, sia usi metaforici e locuzioni, ovvero variazioni tra loro eterogenee.

Ai nostri fini, la distinzione tra variazioni che identificano tipologie azionali e variazioni d’uso di vario tipo è cruciale. Nelle tipologie azionali individuate in figura 1, nonostante la diversità dei diversi tipi, un parlante competente può riconoscere che il verbo è sempre applicato “nel suo significato proprio”, mentre al contrario, negli usi in 4, 6, 7, 9 e 10 il verbo è usato in un significato diverso, ovvero non è riconoscibile in quel contesto un suo significato proprio.

Tale giudizio sintetico di competenza ha conseguenze operative chiare. Nei contesti in cui il verbo è usato nel suo significato proprio, un parlante competente può indicare a qualcuno che non conosce la sua lingua il contesto come una istanza prototipica del concetto (questa e simili cose sono ciò che intendo per ‘aprire’)<sup>3</sup>, innescando quindi un processo produttivo di estensione del concetto alle istanze simili. Questo è il caso in tutti gli usi in figura 1, nonostante questi siano diversi tra loro. Questo non è il caso per gli usi in 4, 6, 7, 9 e 10. Per esempio nessuno potrebbe indicare ad un altro un giocatore che muove una pedina e dire a qualcuno che non conosce la propria lingua “questa e simili cose sono ciò che intendiamo in inglese per *to open*”. Da ciò consegue che l’applicazione del verbo a quel tipo di evento non è libera, ma sottostà a restrizioni idiosincratice che non sono comprese nella capacità di applicare creativamente una espressione.

Chiamiamo *Variazione primaria* di un verbo generale l’insieme dei tipi nei quali il verbo si estende produttivamente sulla base della sua competenza semantica, mentre chiamiamo *Variazione marcata* tutti gli usi che non hanno questa proprietà. La mancata distinzione tra questi tipi di variazione ha come conseguenza che l’ontologia non può predire i contesti nei quali l’applicazione del verbo è produttiva in tutte le istanze. Come vedremo, questa limitazione è assai severa se l’informazione semantica deve essere utilizzata per disambiguare e/o per tradurre un contesto linguistico.

---

<sup>3</sup> Questa perifrasi è la risposta suggerita in Wittgenstein (1953) alla domanda «come potresti spiegare a qualcuno cosa è ‘gioco’ e cosa non lo è?»

### 3. Verbi di azione “di attività”

3.1 Non tutti i verbi che si riferiscono ad azioni sono generali. I verbi chiamati *Activity* nella tradizione (Vendler 1967; Dowty 1979; McCawley 1976) come ‘mangiare’, ‘correre’, ‘stirare’, ecc., si riferiscono, da un punto di vista cognitivo, ad un solo tipo di azione. Si consideri per esempio ‘mangiare’:

- (7a) mangiare la bistecca
- (7b) mangiare il gelato
- (7c) mangiare la minestra



Figura 2. ‘Mangiare’

Da un punto di vista oggettivo le azioni che si realizzano in (7a), (7b) e (7c), hanno differenze evidenti. In (7a) sono implicati i denti, in (7b) la lingua, in (7c) l’atto assomiglia al bere.

Nonostante ciò, da un punto di vista cognitivo non sembra esistere alcun interesse a identificare ciascuno di essi come un tipo a sé stante: in tutti i casi, sia linguisticamente che concettualmente si esegue sempre la stessa azione: ‘mangiare’, ovvero la forma dell’evento non cambia al variare delle sue proprietà e le variazioni si configurano come variazioni di produttività del concetto riscontrabili sullo stesso asse. In altri termini si può sostenere che vari campi di variazione non danno origine a tipi con una variazione interna: non c’è una variazione interna al tipo di cui ‘mangiare il gelato’ è il prototipo.

Questo giudizio correla con il fenomeno del contrasto centro vs. periferia nei concetti naturali. La variazione del verbo risulta *orizzontale*, in quanto disegna un centro e una periferia: ‘mangiare la bistecca’ è un mangiare più prototipico e centrale come istanza del concetto di mangiare rispetto a ‘mangiare il brodo’.

Dunque, diversamente dal caso di ‘aprire’, nel caso di ‘mangiare’ la categorizzazione linguistica e la categorizzazione concettuale vanno, come si dice, la mano nella mano. Per cui, a fronte del termine *generali* per i verbi di azione che si comportano come ‘aprire’, utilizzeremo la qualifica *di attività (Activity)* per indicare i verbi che non si estendono su tipi di azioni diverse (Moneglia 1998).

3.2 Le due tipologie semantiche dei verbi di azione hanno correlazioni diverse a livello delle loro proprietà lessicali e cognitive.

I verbi di attività, come la tradizione che fa capo a Vendler ha evidenziato, correlano con la categoria aspettuale dei *processi* (predicati *atelici*), mentre al contrario i verbi generali correlano massivamente con la categoria degli *eventi* (predicati *telici*). In particolare i verbi di attività non incorrono nel paradosso dell'imperfettività, mentre al contrario vi incorrono i generali:

- “Mario sta mangiando”            *implica*            “Mario ha mangiato già qualcosa”
- “Mario sta aprendo qualcosa” *non implica*    “Mario ha già aperto qualcosa”

I verbi di attività possono essere assoluti, ovvero possono essere saturati anche senza l'oggetto tematico, mentre i verbi generali necessitano della saturazione tematica per dare luogo ad una proposizione:

- Mario mangia
- \*Mario apre

I verbi di attività proiettano immagini mentali in quanto lemmi in isolamento, mentre i verbi generali proiettano immagini mentali solo in connessione alla saturazione tematica.

Sebbene ogni lingua abbia predicati di azione di entrambi i tipi, il loro rapporto quantitativo può cambiare. La maggioranza delle lingue sembra avere la tendenza a registrare più verbi generali, mentre alcune (come il coreano, il danese e il tedesco) sembrano avere la tendenza opposta (Choi & Bowerman 1991; Korzen 2005).

#### 4. Correlazioni cross-linguistiche dei verbi di azione

4.1 La presenza di verbi generali nelle lingue crea problemi seri nella traduzione: non è possibile stabilire nessuna relazione esatta di traduzione tra verbi generali di lingue diverse, ovvero non esiste corrispondenza tra i tipi di azione che sono nell'estensione di verbi generali in lingue diverse. Al contrario i verbi di attività sono quasi sempre in relazione di traduzione pressoché diretta.

Per esempio, data una variazione pragmatica come quella osservata in ‘mangiare’, se una lingua categorizza quel tipo di azione con un verbo di attività, l'applicazione del predicato corrispondente in più lingue risulterà facilmente parallela in tutta la variazione pragmatica del concetto. Per esempio in tutte le lingue romanze e in inglese si utilizza dello stesso verbo (*to eat, manger, comer*, ecc.) in tutti i casi e emergono solo alcune restrizioni d'uso. Per esempio in spagnolo non si

dice *comer sopa* e l'uso richiede piuttosto *haber sopa*, sebbene la restrizione sia sostanzialmente locale.

Non è così nel caso dei verbi generali. Il caso di 'aprire' è stato studiato contrastivamente in dettaglio in per l'inglese e il coreano (Choi & Bowerman 1991):

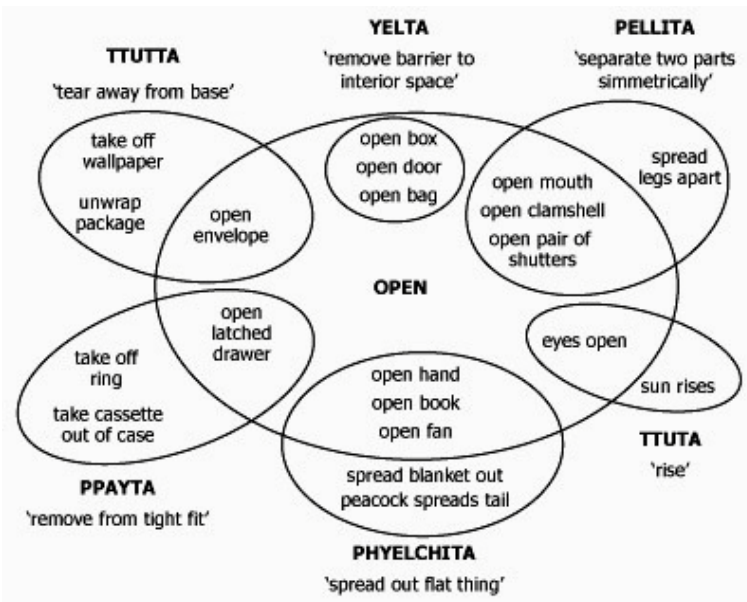


Figura 3. Variazione di *to open* contrastivamente al coreano (da Choi e Bowerman 1991)

I verbi generali rendono problematico il riferimento interlinguistico alle azioni anche più comuni. Per esempio si consideri la variazione del verbo generale italiano ad altissima frequenza 'prendere'. La figura 4 illustra tale variazione attraverso campi identificati da una figurina che istanzia l'azione prototipale. La variazione corrisponde a circa 11 tipi diversi di azione (cfr. Moneglia & Panunzi *submitted*).





	Italian	Spanish	French	English
1.	 <ul style="list-style-type: none"> <li>- prendere il bicchiere</li> <li>- prendere i piatti</li> <li>- prendere la candela</li> <li>- prendere il cappotto</li> </ul>	<ul style="list-style-type: none"> <li>- coger el vaso</li> <li>- coger los platos</li> <li>- coger el candil</li> <li>- coger el abrigo</li> </ul>	<ul style="list-style-type: none"> <li>- prendre le verre</li> <li>- prendre les plats</li> <li>- prendre la bougie</li> <li>- prendre le manteau</li> </ul>	<ul style="list-style-type: none"> <li>- take the glass</li> <li>- take the dishes</li> <li>- take the candle</li> <li>- take the coat</li> </ul>
2.	 <ul style="list-style-type: none"> <li>- prendere l'ombrello per uscire</li> <li>- prendere gli spiccioli</li> <li>- prendere delle cose da mangiare per il viaggio</li> <li>- prendere il cavallo per partire</li> </ul>	<ul style="list-style-type: none"> <li>- coger el paraguas al salir</li> <li>- coger el dinero</li> <li>- coger comida para el viaje</li> <li>- coger el caballo para salir</li> </ul>	<ul style="list-style-type: none"> <li>- prendre le parapluie pour sortir</li> <li>- prendre l'argent</li> <li>- prendre des choses à manger pour le voyage</li> </ul>	<ul style="list-style-type: none"> <li>- take the umbrella and leave</li> <li>- take the money and leave</li> <li>- take the foodstuffs for the journey</li> <li>- take the horse outside</li> </ul>
3.	 <ul style="list-style-type: none"> <li>- prendere il portatile dalla borsa</li> <li>- prendere il liquido dalla provetta</li> <li>- prendere il latte dal freezer per consegnarlo al lattaiolo</li> </ul>	<ul style="list-style-type: none"> <li>- coger (sacar) el portatil de la funda</li> <li>- coger (sacar) el liquido de proveta</li> <li>- coger la leche del congelador para darsela al lechero</li> </ul>	<ul style="list-style-type: none"> <li>- prendre l'ordinateur portable du sac</li> <li>- prendre le liquide de l'éprouvette</li> <li>- prendre le lait du réfrigérateur pour le consigner</li> </ul>	<ul style="list-style-type: none"> <li>- take the laptop out of the bag</li> <li>- take the liquid out of the bottle</li> <li>- take out the milk from the freezer to give to the mikman</li> </ul>
4.	 <ul style="list-style-type: none"> <li>- prendere il libro dall'assistente</li> </ul>	<ul style="list-style-type: none"> <li>- coger el libro del asistente</li> </ul>	<ul style="list-style-type: none"> <li>- prendre (recevoir) le livre à l'assistant</li> </ul>	<ul style="list-style-type: none"> <li>- take the book from the assistant</li> </ul>
5.	 <ul style="list-style-type: none"> <li>- prendere il portafogli alla ragazza</li> <li>- prendere il capello alla signora</li> <li>- prendere i prodotti della terra alla gente del posto</li> </ul>	<ul style="list-style-type: none"> <li>- coger (quitar) el monedero a la chica</li> <li>- coger (quitar) el sombrero a la senora</li> <li>- coger (tomar) los productos de la tierra a los lugareños</li> </ul>	<ul style="list-style-type: none"> <li>- prendre le portefeuille à la fille</li> <li>- prendre le chapeau à la fille</li> <li>- prendre les produits de la terre aux paysans</li> </ul>	<ul style="list-style-type: none"> <li>- take the money from the girl</li> <li>- take the hat off the lady</li> <li>- take the food from the people</li> </ul>

Figura 4a. Variazione di 'prendere' (prima parte)

	Italian	Spanish	French	English
6.	 <ul style="list-style-type: none"> <li>- <i>prendere</i> l'acqua dal rubinetto</li> </ul>	<ul style="list-style-type: none"> <li>- <i>coger</i> el agua del grifo</li> </ul>	<ul style="list-style-type: none"> <li>- <i>prendre</i> l'eau du robinet</li> </ul>	<ul style="list-style-type: none"> <li>- <i>get</i> the water from the fossot</li> </ul>
7.	 <ul style="list-style-type: none"> <li>- <i>prendere</i> il ghiaccio con le pinze</li> <li>- <i>prendere</i> il bordo della stoffa per fare l'orlo</li> </ul>	<ul style="list-style-type: none"> <li>- <i>coger</i> el hielo con las pinzas</li> <li>- <i>coger</i> el borde de la tela para coserlo</li> </ul>	<ul style="list-style-type: none"> <li>- <i>prendre</i> le morceau de glace avec des pinces</li> <li>- <i>prendre</i> le bord du tissu pour faire l'ourlet</li> </ul>	<ul style="list-style-type: none"> <li>- <i>hold</i> the ice with the tongs</li> <li>- <i>hold</i> the material to take the hem</li> </ul>
8.	 <ul style="list-style-type: none"> <li>- <i>prendere</i> il fiore dal giardino</li> <li>- <i>prendere</i> i pomodori nell'orto</li> <li>- <i>prendere</i> le more</li> </ul>	<ul style="list-style-type: none"> <li>- <i>coger</i> la flor del jardín</li> <li>- <i>coger</i> los tomates del huerto</li> <li>- <i>coger</i> moras</li> </ul>	<ul style="list-style-type: none"> <li>- <i>cueillir</i> des fleurs du jardin</li> <li>- <i>cueillir</i> (<i>prendre</i>) les tomates dans le potager</li> <li>- <i>cueillir</i> des mûres</li> </ul>	<ul style="list-style-type: none"> <li>- <i>pick</i> the flower in the garden</li> <li>- <i>pick</i> the tomatoes in the garden</li> <li>- <i>pick</i> the blackberries off the plant</li> </ul>
9.	 <ul style="list-style-type: none"> <li>- <i>prendere</i> il martello a Gina</li> <li>- <i>prendere</i> la spesa alla nonna</li> <li>- <i>prendere</i> le valigie per il cliente</li> <li>- <i>prendere</i> la scatola del tè per qualcuno</li> </ul>	<ul style="list-style-type: none"> <li>- <i>coger</i> el martillo a Gina</li> <li>- <i>coger</i> la compra a la abuela</li> <li>- <i>coger</i> las maletas para el cliente</li> <li>- <i>coger</i> la caja del té para alguien</li> </ul>	<ul style="list-style-type: none"> <li>- <i>apporter</i> le marteau à Gina</li> <li>- <i>apporter</i> les courses à grand-mère</li> <li>- <i>apporter</i> les valises au client</li> <li>- <i>apporter</i> la boîte à thé à quelqu'un</li> </ul>	<ul style="list-style-type: none"> <li>- <i>take</i> the hammer to Gina</li> <li>- <i>take</i> the groceries to grandmother</li> <li>- <i>take</i> the suitcases for the customer</li> <li>- <i>take</i> the tea box for someone</li> </ul>
10.	 <ul style="list-style-type: none"> <li>- <i>prendere</i> il ladro che scappa</li> <li>- il cane <i>prende</i> il biscotto lanciato dal padrone</li> <li>- <i>prendere</i> la biscia nell'acqua</li> </ul>	<ul style="list-style-type: none"> <li>- <i>coger</i> (<i>atrapar</i>) al ladrón que escapa</li> <li>- el perro <i>coge</i> la galleta lanzada por el dueño</li> <li>- <i>coger</i> la serpiente del agua</li> </ul>	<ul style="list-style-type: none"> <li>- <i>attraper</i> le voleur qui s'enfuit</li> <li>- le chien <i>attrape</i> le biscuit lancé par le maître</li> <li>- <i>attraper</i> le serpent dans l'eau</li> </ul>	<ul style="list-style-type: none"> <li>- <i>catch</i> the thief that is fleeing</li> <li>- the dog <i>catches</i> the thrown biscuit</li> <li>- <i>catch</i> the snake in the river</li> </ul>
11.	 <ul style="list-style-type: none"> <li>- <i>prendere</i> la bottiglia con la pallina</li> <li>- <i>prendere</i> il pesce con la fiocina</li> <li>- <i>prendere</i> il nemico alla gamba</li> <li>- <i>prendere</i> lo spartitraffico con la ruota</li> </ul>	<ul style="list-style-type: none"> <li>- <i>golpear</i> la botella con la bola de papel</li> <li>- <i>golpear</i> el pez con el tridente</li> <li>- <i>golpear</i> al enemigo en la pierna</li> <li>- <i>golpear</i> el vaden con la rueda</li> </ul>	<ul style="list-style-type: none"> <li>- <i>toucher</i> la bouteille avec la boule de papier</li> <li>- <i>toucher</i> le poisson avec le harpon</li> <li>- <i>toucher</i> l'ennemi à la jambe</li> <li>- <i>toucher</i> la barrière avec la roue</li> </ul>	<ul style="list-style-type: none"> <li>- <i>hit</i> the bottle with the ball</li> <li>- <i>hit</i> (<i>catch</i>) the fish with the spear</li> <li>- <i>hit</i> the enemy in the leg</li> <li>- <i>hit</i> the guard rail with the wheel</li> </ul>

Figura 4b. Variazione di 'prendere' (seconda parte)



I verbi che in inglese, francese e spagnolo, sono comunemente considerati in relazione di traduzione con ‘prendere’, e in traduzione tra di loro nei dizionari bilingui, sono rispettivamente *coger*, *prendre* e *to take*. Una volta che sia resa esplicita la tipologia di azioni diverse nell’estensione di ‘prendere’ è facile osservare che le estensioni dei predicati inglese, francese e spagnolo non sono in corrispondenza biunivoca. In particolare notiamo che:

- i quattro verbi sono generali perché tutti estendono a eventi qualitativamente diversi;
- nonostante siano, in teoria, in relazione di traduzione, la loro applicazione è parallela solo a un sottoinsieme dei tipi sui quali variano.

In particolare i verbi in questione covariano nei tipi 1-5 di figura 4a, ma non covariano nei tipi di 4b. In particolare per coprire i tipi di ‘prendere’ in 7, 10 e 11, l’inglese necessita rispettivamente di *to hold*, *to catch* e *to hit*, il francese necessita di *attrapper* e *toucher* in 10 e 11, lo spagnolo di *golpear* in 11 e in nessun caso *to take*, *prendre* e *coger* possono estendersi a tali tipi. Dato che non sono equivalenti rispetto alle loro possibili estensioni e quindi non equivalenti in intensione, non può essere stabilita una corrispondenza uno a uno tra tali predicati.

Se poi consideriamo la variazione di ciascuno dei predicati generali considerati in ogni lingua la situazione si complica ulteriormente, in quanto, come è ovvio, l’estensione dei verbi delle altre lingue non è necessariamente un sottoinsieme delle estensioni di ‘prendere’ in italiano. In ogni lingua i verbi generali presentano linee di variazione diverse che non sono possibili estensioni per il verbo generale italiano ‘prendere’. Si veda la figura 5.



a) he takes the book to her

b) she brings the book

a') porta il libro a lei

b') porta il libro

Figura 5. ‘Portare’ vs. *to take* / *to bring*

L'inglese, ad esempio, applica *to take* a molte azioni che in italiano sarebbero piuttosto nell'estensione del verbo 'portare', come in figura 5a, e parallelamente molte azioni che sono 'portare' in italiano non potrebbero essere *to take* in inglese, come figura 5b, che richiede *to bring*.

4.2 Dunque non c'è una relazione di traduzione uno a uno tra verbi generali di azione, ma osserviamo piuttosto una intersezione complessa tra estensioni di predicati generali diversi. Ma, crucialmente, la relazione è sistematica! È di fondamentale importanza per la teoria semantica distinguere tali variazioni dalla fraseologia e dagli usi linguistici idiosincratici che, in quanto tali, sono fatti idiosincratici che possono essere solo annotati a margine delle traduzioni. Nel caso della traduzione dei verbi generali siamo di fronte ad aspetti produttivi della semantica dei verbi in considerazione e la competenza semantica dei parlanti è capace di riconoscere tale variazione in modo creativo per tutte le istanze possibili di ogni tipo. La traduzione dovrebbe quindi essere in linea di principio predicibile.

Il precedente insieme di corrispondenze è una sorgente estremamente ricca di informazioni per quanto riguarda la relazione tra lemmi e tipi azionali. In particolare:

- ogni verbo si applica nel suo significato proprio a vari tipi azionali e nessuno di questi può essere considerato più appropriato dell'altro nella caratterizzazione del suo significato (nessuno è più prototipico dell'altro);
- i verbi sono produttivi in ogni tipo e anche la relazione di traduzione appare produttiva in ogni tipo.

In altri termini se la relazione di traduzione tra verbi generali non è predicibile *across types*, nella variazione verticale di un verbo, al contrario questa è predicibile *within types*, all'interno della sua variazione orizzontale.

Per esempio, se si considerano inglese e italiano, dato che *to take* e 'prendere' sono in relazione di traduzione nel tipo 4.1 e *to take* e 'portare' sono in relazione di traduzione nel tipo 5a, si può prevedere che la relazione di traduzione sussista in tutte le istanze di ogni tipo, indipendentemente dall'oggetto che è coinvolto nell'azione:

- l'attore *takes* / prende un libro / una penna / una tazza / una bambola / dallo scaffale / il mobiletto / la tavola
- l'attore *takes* / porta un libro / una penna / una tazza / una bambola a Mary

La produttività della relazione di traduzione è conseguenza della produttività del tipo, ovvero della sua natura di oggetto semantico.

Lo stesso è vero, in teoria, ovvero fatte salve restrizioni d'uso locali, per ogni tipo. In particolare, le frasi riportate per ogni tipo in figura 4 corrispondono a istanze del verbo riscontrate nel corpus di parlato italiano in C-ORAL-ROM (Cresti & Moneglia 2005) e corrispondono quindi a frasi reali. Si può osservare che in effetti le traduzioni sono sistematiche.

Quindi la messa in rilievo della variazione tipologica di estensione dei predicati generali evidenzia un aspetto creativo e sistematico della competenza semantica di ogni lingua. La sistematicità delle traduzioni all'interno del tipo conferma che i tipi azionali possono essere considerati un livello ontologico indipendente dal linguaggio.

La produttività della traduzione all'interno dei tipi azionali nelle lingue naturali è un risultato fondamentale per la teoria della traduzione e più in generale per l'oggettivazione del modello della cognizione umana. Nonostante l'ipotesi Sapir-Whorf, la produttività della traduzione all'interno delle istanze di un tipo è una prova che categorizziamo le azioni nello stesso modo, indipendentemente dalla nostra lingua, lingua che mostra invece modalità idiosincratiche di categorizzazione.

Il motivo per cui non possiamo nei fatti operare questa previsione a livello cross-linguistico è duplice e riguarda sia le nostre conoscenze sull'ontologia dell'azione sia quelle sul modo con cui questa viene interpretata nelle lingue: a) sappiamo poco di quali siano i verbi generali; b) non conosciamo la variazione su tipi azionali a cui i predicati generali si riferiscono (la loro variazione primaria); c) a ben vedere non sappiamo granché dell'ontologia dell'azione nella nostra vita quotidiana, ovvero non abbiamo cognizione di quante e quali siano le "azioni effettive" a cui ci riferiamo con il linguaggio. Ci viene comodo, infatti, identificare le azioni con i verbi di azione della nostra lingua, salvo che tale identificazione è fallace, perché il verbo generale non si riferisce ad un'azione, ma a molte. Nel paragrafo seguente vedremo con quale frequenza e con quali modalità avviene il riferimento all'azione nella lingua d'uso.

## 5. I verbi di azione nei corpora di lingua parlata

Il linguaggio ordinario non rispecchia l'ontologia dell'azione e in particolare i verbi generali si riferiscono in modo produttivo a più tipi di azioni diverse. Inoltre i verbi generali non variano nello stesso modo in lingue diverse, per cui la traduzione di tali predicati non può essere predetta. Ciò è tanto più imbarazzante perché tale variazione non è idiosincratica, ma ha carattere di produttività e, come vedremo, l'uso dei verbi generali è ad alta frequenza nel parlare quotidiano, nel quale il riferimento all'azione è centrale per la comunicazione. Si pone quindi il problema scientifico e infrastrutturale di individuare le variazioni significative dei verbi generali di lingue diverse nell'uso quotidiano.

Il riferimento linguistico alle azioni e la rilevanza che queste hanno nella vita di ogni giorno va di pari passo. Più un'azione è frequente e importante e più sarà frequente il riferimento a questa nei vari contesti del parlato spontaneo. Quindi, in linea di principio, dall'osservazione della variazione d'uso dei verbi di azione nel parlato possiamo indurre sia la variazione primaria dei verbi in considerazione sia il riferimento all'ontologia dell'azione. Per questo motivo i corpora di parlato spontaneo hanno un'importanza cruciale nello studio del rapporto che le diverse lingue hanno con le azioni.

Uno studio pilota (Panunzi & Moneglia 2004) basato sul LIP (De Mauro et al. 1993), ha ipotizzato che i predicati generali siano la classe più frequente nel lessico verbale sia per numero di occorrenze sia per numero di lemmi. I dati qui riportati mostrano che questo risultato si replica a livello cross-linguistico su corpora di ordine superiore. I dati presentati nelle tavole di seguito sono tratti da un corpus di parlato Italiano (1.284.978 tokens), da un corpus orale spagnolo (1.104.627 tokens) e dalla parte orale del British National Corpus (BNC; 10.378.225 tokens), che è stata utilizzata come riferimento normativo per misurare la significatività dei corpora romanzi (Moneglia & Panunzi 2007)<sup>4</sup>.

Nonostante il fatto che la dimensione del BNC sia di ordine superiore, il lessico fondamentale delle risorse si è rivelato strettamente comparabile. Rispettivamente 1.588 entrate lessicali contenenti 287 verbi nella risorsa inglese, 1.590 entrate di cui 299 verbi nella risorsa italiana e 1.303 entrate di cui 314 verbi nella risorsa spagnola<sup>5</sup>. In altri termini, il passaggio da 1 a 10 milioni di parole non sembra ripercuotersi sulla rappresentazione del lessico di base nei corpora. Aggiungendo l'insieme addizionale sono state generate, per ogni lingua, liste di frequenza verbali, di 405 entrate ciascuna, che rappresentano il lessico verbale a maggior probabilità d'uso. In tali liste di frequenza verbali la porzione di lessico specificamente orientata all'identificazione delle azioni è stata separata da quella che non si riferisce ad azioni. In particolare il lessico verbale che introduce subordinate (*dicendi, sentiendi, putandi, performativi, circostanziali* ecc.) per definizione non si riferisce ad azioni, ed è stato considerato come un insieme separato.

I lessici sono stati considerati sia in termini di lemmi che in termini del numero di occorrenze di questi nel corpus. I dati mostrano una stretta comparabilità a livello cross-linguistico dell'uso linguistico verbale. Il numero di lemmi che si riferiscono ad azioni è maggiore rispetto a quelli che svolgono funzione subordinante (con un fattore di 5/3). Dal punto di vista delle occorrenze nel corpus di tali lemmi, queste si

---

<sup>4</sup> Il corpus Italiano utilizzato è una collezione dei corpora: C-ORAL-ROM Italian, LABLITA e LIP. Il Corpus Spagnolo raccoglie C-ORAL-ROM Spanish e CORLEC. Sia queste risorse sia il BNC orale sono stati lemmatizzati e PoS taggati con Treetagger.

<sup>5</sup> Le occorrenze dei modali, degli ausiliari e delle copule non sono state considerate nelle statistiche. Tali occorrenze corrispondono a circa il 50% delle occorrenze verbali nel parlato spontaneo.

distribuiscono equamente tra i due tipi. Per cui possiamo osservare che nel parlato spontaneo circa la metà delle occasioni in cui si usa un verbo, questo si riferisce ad un'azione.

Per una analisi più qualitativa, i verbi orientati all'azione sono stati divisi in classi in modo tale che fosse possibile isolare quelle voci che si riferiscono all'orizzonte azionale primario e le occorrenze relative ad azioni di base sono state distinte dalle occorrenze relative ad azioni di livello superiore.

A tal fine, tutti i verbi che per il loro contenuto semantico non si riferiscono alla sfera primaria dell'azione sono stati separati in un insieme spurio (OTHZ) comprendente: (a) verbi di attività sociali ('comprare', 'affittare', 'dare il benvenuto' ecc.); (b) verbi di attività che si riferiscono allo scambio dialogico ('citare', 'discutere', 'replicare'); (c) verbi di attività di tipo astratto e che non puntano su tipologie azionali specifiche ('modificare', 'derivare', 'migliorare', 'aumentare'; ecc.).

I verbi rimasti sono stati considerati a priori riguardanti azioni primarie e sono stati divisi in tre classi: (a) generali (GEN); (b) di attività (ACT); (c) di movimento (MOV).

I dati mostrano che i verbi orientati all'azione di base (MOV, ACT e GEN) coprono la maggioranza delle occorrenze che designano azioni. Inoltre, questo gruppo di verbi è costituito da un gruppo di lemmi molto ridotto (170 per l'inglese, 158 per l'italiano, 130 nello spagnolo). Inoltre all'interno di questo gruppo le tre lingue sono caratterizzate da un lessico orientato ai predicati generali. La classe GEN comprende infatti il doppio dei verbi di attività e cinque volte il numero dei lemmi dei verbi di movimento.

Italiano, spagnolo e inglese sono quindi simili per quanto riguarda la struttura del lessico verbale fondamentale e le modalità con cui questo è proiettato nell'uso linguistico. Il numero di verbi usati per riferirsi ad azioni di base si dimostra quindi ristretto a livello cross-linguistico, ma la loro incidenza nell'uso è alta. La probabilità di riferirsi ad una azione attraverso un verbo generale, nelle lingue prese in considerazione, è da 3 a 5 volte più alta che attraverso un verbo di attività e circa 2 volte più alta rispetto al movimento. Il lessico verbale dell'italiano, dello spagnolo, in modo simile all'inglese si caratterizza quindi specificamente per il riferimento all'azione attraverso verbi generali e prevede quindi un'ampia variazione nella tipologia delle azioni individuate nell'uso.

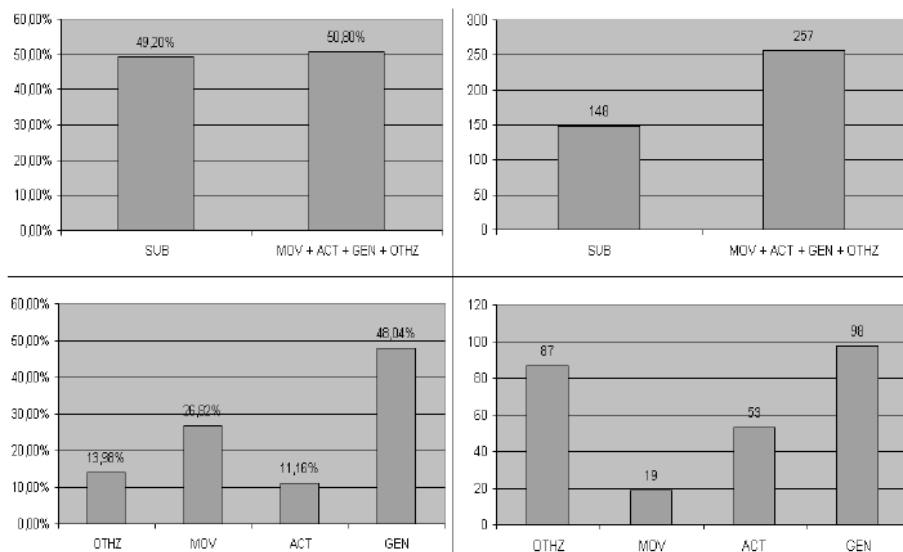


Figura 6. Percentuale di occorrenze e numero di lemmi nel lessico verbale ad alta frequenza per tipi (inglese parlato)

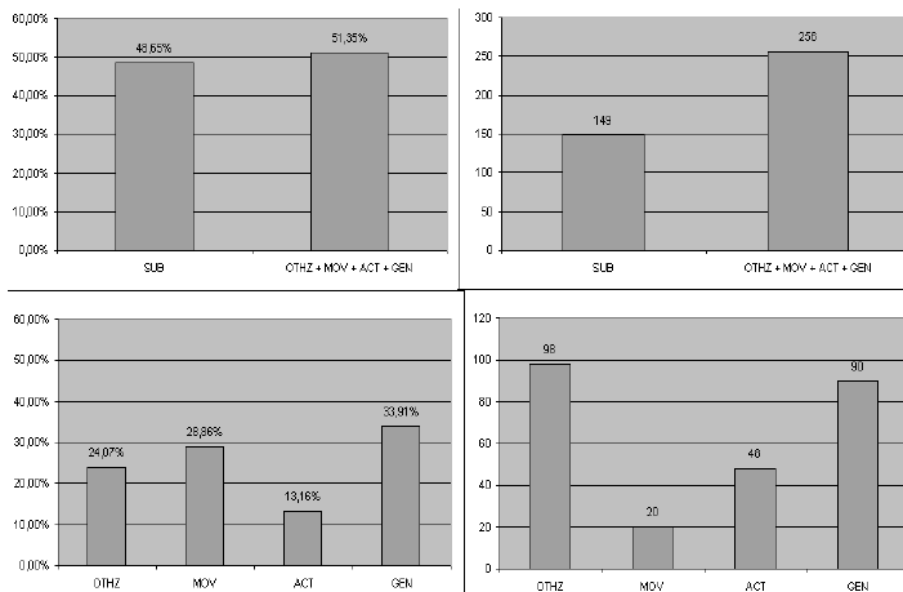


Figura 7. Percentuale di occorrenze e numero di lemmi nel lessico verbale ad alta frequenza per tipi (italiano parlato)

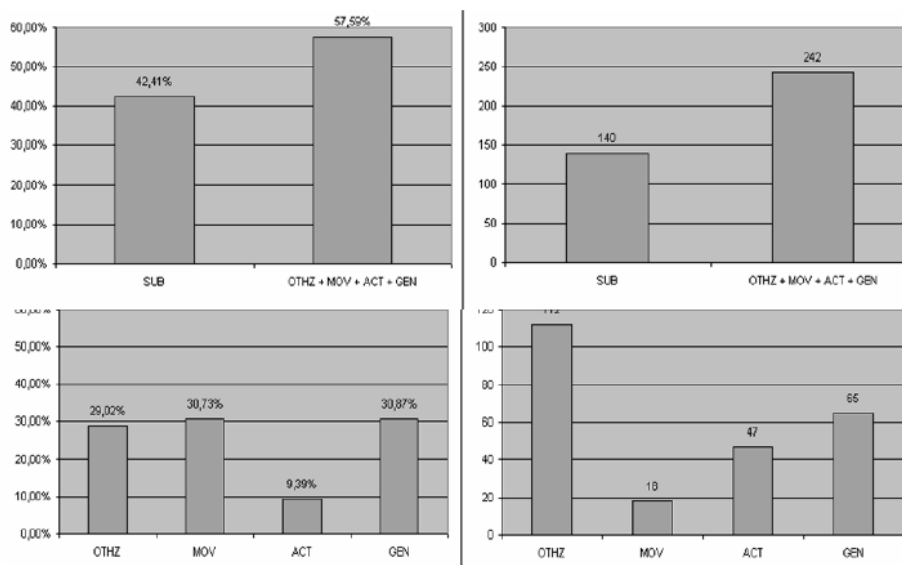


Figura 8. Percentuale di occorrenze e numero di lemmi nel lessico verbale ad alta frequenza per tipi (spagnolo parlato)

## 6. Conclusioni

I verbi generali fanno parte del lessico ad alta frequenza nel parlato spontaneo e ciò rende più sensibile il problema della traduzione anche dal punto di vista dell'incidenza quantitativa. Il numero dei verbi generali nel lessico ad alta frequenza è ridotto, ma la loro incidenza nell'uso è massiccia. I corpora di parlato, nei quali il riferimento all'azione si manifesta nell'uso linguistico, mostrano una stretta regolarità nelle modalità di riferimento verbale all'azione, sia qualitativa che quantitativa.

L'informazione riguardante la relazione tra azioni in quanto entità ontologiche e la loro codifica lessicale cross-linguistica può dunque essere derivata da risorse già esistenti. I corpora di parlato spontaneo contengono infatti un riferimento alle azioni più frequenti nella vita di ogni giorno e, insieme, anche la loro codifica lessicale. Le azioni più frequenti possono dunque essere identificate in parallelo ai verbi di azione ad alta frequenza nel parlato.

Il progetto IMAGACT intende realizzare un'infrastruttura che può ridurre fortemente i limiti attuali dei sistemi di *Natural Language Understanding* e *Automatic Translation* per quanto riguarda il riferimento all'azione, realizzando una Ontologia dell'Azione Interlinguistica che specifica in modo non ambiguo l'ambito di variazione dei verbi di azione in lingue diverse (italiano, inglese) e ne permette l'intertraducibilità.

IMAGACT utilizzerà metodologie sia *corpus based* che *competence based* per estrarre sia i tipi ontologici azionali sia la loro codifica linguistica da corpora italiani, inglesi e francesi. L'insieme dei dati varierà tra i 5 e i 6 milioni di parole (per un minimo di 1,5 MW per lingua). Sulla base degli studi preliminari, il corpus permetterà l'induzione di circa 4000 tipi di azione ad alta frequenza mappati su circa 200 *synsets* lessicali per lingua. I *synsets* saranno integrati da informazione semantica, sintattica e statistica derivata dai corpora stessi.

Il progetto produrrà un'Ontologia Interlinguistica dell'Azione sviluppata in una base di dati e permetterà l'associazione ad ogni tipo ontologico del predicato corrispondente in ogni lingua considerata (italiano, inglese).

## Bibliografia

- Bruner, J. 1983. *Child's talk. Learning to use language*. New York: Norton.
- BNC. <http://www.natcorp.ox.ac.uk/>
- Choi, S. & Bowerman, M. 1991. Learning to express motion events in English and Korean: the influence of language specific lexicalization patterns. *Cognition* 41: 83-121.
- CORLEC. <http://www.llf.uam.es/ESP/Corlec.html>
- Cresti, E. & Moneglia, M. (eds) 2005. *C-ORAL-ROM Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: Benjamins.
- De Mauro, T., Mancini, F., Vedovelli, M. & Voghera, M. 1993. *Lessico di frequenza dell'italiano parlato*. Milano: ETASLibri.
- Dowty, D. 1979. *Word meaning and montage grammar*. Dordrecht: Reidel.
- Givon, T. 1986. Prototypes: Between Plato and Wittgenstein. In C. Craig (ed), *Noun Classes and Categorization*. Amsterdam: Benjamins, 77-102.
- IMAGACT. <http://lablita.dit.unifi.it/projects/IMAGACT>
- Korzen, I. 2005. Endocentric and esocentric languages in translation. *Perspectives: Studies in translatology* 13(1): 21-37.
- LABLITA <http://lablita.dit.unifi.it/corpora/>
- McCawley, J. 1976. *Grammar and Meaning: Papers on Syntactic and Semantic Topics*. New York: Academic Press.
- Minsky, M. 1968. *Semantic Information Processing*. Cambridge (MA): MIT Press.
- Moneglia, M. 1998. Teoria empirica del senso e partizione semantica del lessico. *Studi di Grammatica Italiana* XVII: 363-398.
- Moneglia, M. & Panunzi, A. 2007. Action Predicates and the Ontology of Action across Spoken Language Corpora. The Basic Issue of the SEMACT Project. In M. Alcántara, T. Declerck (eds), *Proceeding of the International Workshop on the Semantic Representation of Spoken Language*. Salamanca: Universidad de Salamanca, 51-58.
- Moneglia, M. & Panunzi, A. (submitted). General verbs, Activity verbs and natural language ontology. The cross-linguistic productivity of action types in spontaneous speech corpora. Paper presented at the Conference Verb Typologies Revisited: A Cross-



- linguistic Reflection on Verbs and Verb Classes, February 5-7 2009, Ghent, Belgium.  
Submitted for publication in *Languages in Contrast*.
- Panunzi, A. & Moneglia, M. 2004. La variazione primaria del verbo nel lessico dei corpora di parlato. In F. Albano Leoni, F. Cotugno, M. Pettorino, R. Savy (eds), *Atti del convegno nazionale "Il parlato Italiano"*. Napoli: M. D'Auria Editore, 1-24 (C4).
- Piaget, J. 1954. *The Construction of Reality in the Child*. New York: Basic Books.
- Rosch, E. 1978. Principles of categorization. In E. Rosch, B.B Lloyd (eds) *Cognition and categorization*. Hillsdale (NJ): LEA, 27-48.
- Tomasello, M. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge (MA): Harvard University Press.
- TreeTagger. <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>
- Vendler, Z. 1967. *Linguistics in Philosophy*. Ithaca: Cornell University Press.
- Vocabolario degli Accademici della Crusca* 1612. Venezia, Giovanni Alberti.
- Von Wright, G.H. 1963. *Norm and Action: A Logical Enquiry*. London: Routledge & Keegan.
- Wittgenstein, L. 1953. *Philosophical Investigations*. Oxford: Blackwell.
- WordNet. <http://wordnet.princeton.edu/>



# ITALIANO LINGUA ESOCENTRICA: L'USO DEI TEMPI VERBALI NELLA NARRAZIONE

Luisa Amenta  
Università di Palermo

## 1. Premessa

Nell'ambito della distinzione tipologica tra lingue endocentriche ed esocentriche e del rapporto tra schemi cognitivi e lingua, con il presente contributo ci proponiamo di indagare l'uso dei tempi verbali nella narrazione e la conseguente modalità di strutturazione del testo in italiano, al fine di avere ulteriori elementi circa la tendenza di questa lingua all'esocentricità.

Come è noto, la distinzione tipologica tra lingue endocentriche ed esocentriche è tradizionalmente basata sulla differente distribuzione lessicale. In questa prospettiva, se limitiamo la nostra attenzione all'ambito verbale, le lingue germaniche si caratterizzano per la presenza di verbi più specifici dal punto di vista lessicale ma morfologicamente più poveri mentre i verbi delle lingue romanze sono più astratti e generici dal punto di vista lessicale ma più articolati per ciò che concerne la ricchezza della flessione.

Questa opposizione tra le due tipologie linguistiche si riflette, anche a livello di macrostruttura, sulla modalità di articolazione del testo per cui, seguendo il modello proposto da Korzen (2004; 2005), le lingue endocentriche tendono verso la paratassi e verso una strutturazione testuale più lineare mentre quelle esocentriche, come l'italiano, verso l'ipotassi. In quest'ottica i verbi dell'italiano sono i marcatori di una focalizzazione gerarchica delle relazioni tra gli eventi che ne riflette lo schema cognitivo soggiacente.

In proposito, Korzen (2005: 130) evidenzia la relazione che esiste tra la lingua e lo schema cognitivo che determina la testualizzazione e osserva come la tendenza alla strutturazione lineare delle lingue endocentriche comporti “un'equa distribuzione di attenzione sulle varie azioni o situazioni di una sequenza, in cui ognuna viene vista individualmente e in modo relativamente concreto”. Il

coinvolgimento di poche entità diverse alla volta comporta pure un incoraggiamento al riferimento cronologico dei vari eventi.

Di contro, in lingue come l'italiano la ricchezza flessiva si accompagna a una strutturazione più ipotattica e a una testualizzazione più gerarchica. Come osserva ancora Korzen (2005: 130), la scelta obbligatoria tra più forme verbali “costringe gli italiani a valutare e ad esplicitare la relazione tra le varie proposizioni e quindi a distinguere eventualmente tra livelli pragmatico-narrativi diversi. Da ciò consegue una particolare attenzione verso la relazione tra i vari eventi extralinguistici e quindi un incoraggiamento a considerare nella stessa unità cognitiva, in modo più sintetico, frammenti più grandi di input extralinguistico, frammenti con più parti tra cui viene interpretata e valutata la relazione”.

Ciò come conseguenza del fatto che i vari livelli pragmatico-narrativi, gestiti grazie ai vari tempi del testo, sono articolati proprio in virtù dello stretto legame tra lingua e cognizione, tra narrazione e schema cognitivo/interpretativo della realtà. Sintetizzando, per ciò che concerne l'articolazione dei tempi verbali nel testo lingue endocentriche ed esocentriche si comporterebbero come illustrato nella tabella seguente:

Tabella 1

<b>Lingue endocentriche</b>	<b>Lingue esocentriche</b>
Testualizzazione più parattattica	Testualizzazione più ipotattica
Attenzione per la singola azione	Attenzione per la gerarchia delle azioni
Eventi considerati linearmente	Distinzione di più livelli pragmatico-narrativi
Preciso riferimento cronologico ai singoli eventi	Riferimenti alla relazione tra gli eventi

## 2. Il *corpus* di analisi

Per verificare l'applicabilità della tendenza all'esocentricità dell'italiano, relativamente alla strutturazione delle relazioni temporali in un testo scarsamente pianificato, come quello orale, abbiamo scelto di orientare la nostra indagine ad un *corpus* di parlato<sup>1</sup>. Abbiamo pertanto rivolto la nostra attenzione alle produzioni di

<sup>1</sup> Nella classificazione tipologica delle lingue, in relazione alle modalità di pianificazione narrativa, un ruolo di primo piano è giocato dalla variabile diamesica considerata. In proposito emblematici i risultati contrastivi ottenuti dal gruppo di ricerca sulla lingua scritta e orale del *corpus* di “Mister Bean” in italiano e in danese, per cui si veda Jansen, B.L. Jensen, E.S. Jensen, Korzen, Polito, Skytte & Strudsholm (1997).

parlato italiano a codice bloccato, elicitate nell'ambito delle inchieste sociovariazionali dell'Atlante Linguistico della Sicilia (ALS).

Come osserva Paternostro (2004), a cui si rimanda integralmente per tutti i problemi metodologici legati all'elicitazione di parlato nelle interviste sociovariazionali dell'ALS, la mancanza di libertà nella scelta del codice da parte dell'informatore e l'indicazione di una durata minima del testo caratterizzano prettamente il nostro parlato e lo differenziano rispetto al parlato spontaneo tout court con cui condivide, secondo la definizione che ne dà Cresti (2000), la simultaneità tra pianificazione linguistica e atto di parola.

È un parlato dunque altamente regolato anche perché il locutore si trova a dover produrre un particolare tipo di testo, tendenzialmente monologico, in una situazione di media formalità e con un interlocutore, il raccoglitore, che il più delle volte limita i propri interventi a *feedback* e a sollecitazioni o per il mantenimento del tema o per la scelta di un altro tema che permetta all'informatore di parlare di più.

Sebbene nello specifico dell'analisi si prenderanno in considerazione precipuamente i turni prodotti dagli informatori, questi non sono mai stati considerati a prescindere dalla globalità dello scambio interazionale che è avvenuto con il raccoglitore, che viene parimenti considerato un soggetto attivo dell'evento comunicativo, che condiziona più o meno consapevolmente la produzione dell'informatore, secondo quelli che sono i presupposti teorici alla base delle scelte metodologiche operate nell'ambito dell'ALS, per cui si rimanda a D'Agostino & Paternostro (2006).

Il corpus su cui basiamo la nostra indagine è costituito da venti ore complessive di parlato prodotte da diciassette informatori per quattordici punti di inchiesta rappresentativi del territorio siciliano per un totale di 238 testi esaminati.

Per ogni punto, il campione dell'ALS prevede cinque tipologie di gruppi familiari con un ordine triadico "Nonno-Genitore-Figlio" che variano in relazione al grado di istruzione e alla prima lingua (italiano o dialetto) dei componenti, più due adolescenti uno di istruzione alta e uno di istruzione bassa.

I brani esaminati per la presente inchiesta sono stati prodotti da informatori differenziati in base all'età e al livello di istruzione, a cui era stato richiesto di parlare di episodi del proprio passato per cinque minuti in italiano e cinque in dialetto. In questa sede, limiteremo la nostra attenzione ai cinque minuti di parlato in italiano degli informatori della generazione dei nonni e di quella dei figli in modo da poter disporre di un campione volutamente polarizzato per quanto concerne l'età.

Le storie dei nostri informatori si riferiscono sia ad episodi recenti che a racconti di gioventù che ci hanno permesso di verificare l'alternanza dei tempi del passato in base anche al parametro della distanza temporale. Altri parametri che abbiamo preso in considerazione sono il grado di formalità e l'oggettività che il locutore ha voluto conferire al suo racconto.

Altrettanto rilevante nell'analisi dei dati è stata l'attenzione per le dimensioni diastratica e diagenazionale di variazione per verificare se e in che modo queste interagiscono con la tendenza tipologica dell'italiano. In tal senso, si esaminerà se il livello di istruzione, la maggiore o minore dimestichezza con l'italiano e l'età possano comportare differenze nelle modalità di gestione e di articolazione dei piani narrativi.

### 3. Modelli interpretativi testuali e pragma-narrativi

L'analisi dei tempi verbali è una tematica non nuova nel campo della linguistica testuale, ciononostante continua ad essere un nodo centrale per descrivere ed interpretare la complessa rete di relazioni che si instaura all'interno di un testo. Infatti, come è noto e come è stato ampiamente dimostrato da Lo Cascio (1982), Lo Cascio & Vet (1986) e Bertinetto (1986) per l'italiano, una parte della coesione del testo è affidata ai rapporti anaforici e/o deittici che legano i vari tempi verbali tra loro ed in relazione al momento di enunciazione.

Tuttavia, l'articolazione dei tempi in un testo non ha esclusivamente la funzione di far avanzare il piano narrativo e di veicolare informazioni relative alla progressione lineare degli eventi nel tempo, secondo un modello iconico di rispecchiamento del loro susseguirsi reale, ma ha anche quella di segnalare quello che Weinrich (1978) definisce l'"atteggiamento comunicativo" del parlante nei confronti del suo enunciato, dal momento che questi opera una pianificazione del testo che prevede una selezione dei vari tempi in virtù di un suo maggiore o minore coinvolgimento.

L'analisi dei tempi nel testo permette quindi di esplicitare l'articolazione del contenuto informativo in relazione alla rilevanza che gli eventi assumono per il parlante, operando una distinzione tra informazioni di *foreground* vs. *background*: primo piano vs. sfondo narrativo che trova in italiano una sua chiara applicazione nell'opposizione tra i passati remoto e prossimo da una parte e imperfetto e trapassato dall'altra. In tal senso, se l'alternanza tra tempi imperfettivi e perfettivi consente di separare informazioni di *background* da quelle di *foreground*, più difficile è analizzare l'alternanza tra tempi che sembrano possedere gli stessi tratti aspettuali. E inoltre, se ci spostiamo sul piano del confronto interlinguistico, come osserva ancora Korzen (2005), l'esplicitazione della distinzione tra *foreground* e *background* pragma-narrativi risulta essere di più difficile realizzazione nelle lingue endocentriche rispetto a quelle esocentriche per la minore ricchezza morfologica e flessiva delle prime rispetto alle seconde.

Se consideriamo i modelli teorici che tradizionalmente si sono imposti nel campo della linguistica testuale, per ciò che concerne l'articolazione dei tempi, Benveniste (1971), Weinrich (1964), Lo Cascio (1982) e Lo Cascio & Vet (1986):

Tabella 2

<b>Benveniste (1959)</b>	<b>Weinrich (1964)</b>	<b>Lo Cascio (1982)</b>
<i>Piano della storia:</i>	<i>Tempi narrativi:</i>	<i>Tempi deittici:</i>
<ul style="list-style-type: none"> <li>- narrazione distaccata degli eventi passati;</li> <li>- maggiore oggettività.</li> </ul>	<ul style="list-style-type: none"> <li>- racconto di eventi passati</li> <li>- atteggiamento di distensione</li> <li>- eventi di primo piano</li> </ul>	<ul style="list-style-type: none"> <li>- macroinsieme degli eventi principali della narrazione;</li> <li>- relazione diretta con il momento di riferimento.</li> </ul>
(Passato Remoto, Imperfetto)	(Trapassati, Passato Remoto, Imperfetto)	(Presente, Passato Prossimo, Passato Remoto, Imperfetto, Futuro)
<i>Piano del discorso:</i>	<i>Tempi commentativi:</i>	<i>Tempi anaforici:</i>
<ul style="list-style-type: none"> <li>- enunciazione con emittente e destinatario;</li> <li>- maggiore soggettività</li> </ul>	<ul style="list-style-type: none"> <li>- maggiore coinvolgimento del locutore;</li> <li>- atteggiamento di tensione</li> <li>- sfondo narrativo.</li> </ul>	<ul style="list-style-type: none"> <li>- elementi secondari che riguardano aspetti degli eventi principali;</li> <li>- relazione con i tempi deittici.</li> </ul>
(Presente, Futuro, Passato Prossimo, Imperfetto)	(Passato Prossimo, Presente, Futuro)	(Trapassati, Futuro Ant., Imperfetto)

possiamo osservare come in particolare Weinrich e Lo Cascio assumono il concetto di avanzamento della narrazione come momento centrale nella strategia di pianificazione che attua il locutore e di conseguenza classificano i vari tempi verbali in base alla loro capacità di portare avanti o meno la storia. Il passato remoto è il tempo del primo piano, della messa in rilievo della trama narrativa che fissa gli eventi passati neutralizzandoli e distanziandoli attraverso il filtro della narrazione, l'imperfetto è il tempo dello sfondo, che costituisce la cornice della trama narrativa e il passato prossimo è un tempo commentativo che fissa il passato permettendone l'attualizzazione. Dunque, teoricamente, l'alternanza di questi tempi in un testo dipende dalla strategia perseguita dal narratore, dalla scelta di ciò che vuole mettere in primo piano.

Ovviamente la temporalità nel parlato assume caratteristiche proprie non perfettamente riconducibili alle strategie testuali della narrativa nello scritto, dal momento che avviene una concentrazione su poche forme morfologiche che estendono le loro potenzialità d'uso rispetto alla lingua standard. Inoltre, il peso del sostrato dialettale nelle varietà regionali di italiano può intervenire nel comportare differenze rispetto a quanto si può osservare per lo standard.

In questa prospettiva, verificheremo se il nostro *corpus* di parlato di italiano regionale connotato diastraticamente e diagenetico, permette di scorgere differenze rispetto le tendenze tipologiche della lingua standard per cui si potrebbe





emerge dal testo è una successione lineare di eventi affidata indifferentemente all'uso dei passati prossimi o agli imperfetti narrativi abituali.

Il momento di riferimento e la distanza temporale che intercorrono tra l'evento narrato e il momento di enunciazione non sembrano costituire in effetti un fattore che orienta la scelta degli informatori nell'uso di una tempo verbale piuttosto che un altro, anche se l'evento narrato presenta le caratteristiche di compiutezza e si presenta come interamente concluso in un passato anche lontano, per cui si vedano i frammenti da (2) a (5):

- (2) Io a diciotto anni **MR1** **E1** *mi sono arruolato* sotto le armi. *Ho conseguito / durante il servizio* **E2** **MR2** militare il grado di sergente maggiore. Il mio *era* il settantaseiesimo reggimento **S1** [...]. Io *mi sono accattivato* **E3** la simpatia dell'aiutante maggiore del | della compagnia. (N., istr. alta, Agrigento).
- (3) E' successo una sera, **MR1** per due ragazze, che in discoteca, un mare di confusione a fare e a dire, *ci siamo bisticciati*, **E1** poi *sono venuti* i carabinieri, **E2** *ci hanno fatto* una denuncia. **E3** (Ad., istr. bassa, Trapani).
- (4) Nel 1938 **MR1** *mi hanno chiamato* sotto le armi, **E1** *sono andato* a Savona e poi *ci hanno portato* al distaccamento di Millesimo. Di Millesimo poi *sono andato / a* Bologna. **E2** **E3** **E4** (N., istr. media Agrigento).
- (5) E così *è venuta* una zia dopo una settimana **E1** **MR1** / a casa mia, dicendomi che quella famiglia *che avevamo incontrato* a fare la visita, *aveva* un figlio che *si doveva fare fidanzato*. **S1** (N., istr. media, Trapani).

Negli esempi riportati la distanza temporale viene espressa tramite l'uso dei localizzatori nel passato (*a diciotto anni, una sera, nel 1938*) e non con il ricorso al passato remoto. Inoltre, l'avanzamento della storia da un evento all'altro avviene attraverso l'uso del passato prossimo, contrariamente a quanto previsto dal modello di Weinrich secondo cui l'uso di questo tempo, sebbene permetta di esprimere un maggiore coinvolgimento del locutore, è destinato allo sfondo narrativo.

Il caso del frammento riportato in (5) costituisce una eccezione, rispetto a quanto sin qui osservato, dal momento che è proprio l'uso del trapassato prossimo

permettere di articolare il periodo in più piani narrativi, creando con gli imperfetti, lo sfondo della situazione narrata<sup>3</sup>. Va osservata nello stesso frammento anche la stretta correlazione che si viene a creare tra subordinazione e uso del trapassato prossimo per creare il *backgrounding* narrativo.

Dunque, la tendenza alla riduzione su un piano narrativo di tutta la successione degli eventi attraverso una serie di passati prossimi sembra essere strettamente legata al livello di istruzione. Infatti, se per il momento limitiamo la nostra attenzione alla generazione dei nonni, possiamo osservare che se da una parte i nonni di istruzione medio-alta si distinguono per la produzione di testi narrativi con una sintassi non semplicemente giustappositiva, ma che si arricchisce di più piani che rispecchiano l'alternarsi di informazioni di *foreground* e *background*, dall'altra i nonni di istruzione bassa appiattiscono la linea degli eventi scegliendo prevalentemente l'uso del passato prossimo. In proposito, va notato che ricorrono all'uso del passato prossimo anche persone anziane con un basso livello di istruzione, dialettone come prima lingua, per le quali sarebbe facile ipotizzare una interferenza del sostrato dialettale e dunque una maggiore predisposizione all'impiego del passato remoto. In particolare nel frammento (6) l'informatrice ricorre ad un passato prossimo per un evento lontano pur non padroneggiando la selezione dell'ausiliare e risentendo nel ricorso ad *avere* dell'influsso dialettale:

- (6)  $\underbrace{\text{La vita mia sempre accusi ha statu.}}_{\text{S1}}$  Poi mi ha sposato, a venticinque anni mi ha,  $\text{E1}$   $\text{MR1}$  =E1  
 sposato quando mi ho sposato / mi ho messo tre fratelli dentro.  
 $\text{MR2}$   $\text{E2}$   
 (N., istr. bassa, Agrigento).

Il frammento (7) prodotto da un informatore di istruzione bassa costituisce un ulteriore esempio di come l'alternanza tra i tempi del passato non si possa ricondurre alla distanza temporale quanto piuttosto ad una strategia di pianificazione del testo per cui la salienza degli eventi narrati nella vita dell'informatore comporta la scelta del passato prossimo o del passato remoto. In proposito, si osservino, in particolare, i momenti di riferimento 1 e 2 e gli eventi E1 al passato prossimo per esprimere quello cronologicamente più distante e E2 al passato remoto per quello più vicino. Inoltre, si noti come il passato prossimo si mantiene nel prosieguo della narrazione come il tempo più usato.

<sup>3</sup> Sui valori del Trapassato Prossimo si rimanda a Korzen (2002).

- (7) io all'età di quattordici anni / per ragioni di famiglia / *sono andato* al pastificio  
**MR1** **E1**  
 Mulino San Giovanni / a lavorare la pasta. / all'età di sedici anni e mezzo  
**MR2**  
*me ne andai* volontario / nella Milizia Artiglieria Contraerea e *partii* il dieci / il dieci  
**E2** **E3**  
gennaio del millenovecentoquarantadue / perché l'Italia / *aveva* bisogno / di noi. / mi  
**MR3** **S1**  
*hanno portato* / dopo di svariati paesi / per i corsi che si facevano / a Bologna // e mi  
**E4**  
*hanno assegnato* alla terza sessione / mitragliatrici di sette millimetri e cannoncini di  
**E5**  
 venti millimetri. / giugno del duemila | del mille / dunque / millenovecento // giugno  
del millenovecentoquarantadue / *siamo stati chiamati* in Sicilia per difendere la  
**MR4** **E6**  
 Sicilia/ dove *siamo andati a finire* /a la Piana di Catania.  
**E7**  
 (N., istr. bassa, Sommatino – Caltanissetta)

Nello stesso frammento (7) va osservato anche che lo sfondo narrativo presente nella subordinata causale è espresso tramite l'imperfetto, tempo che viene così a opporsi funzionalmente ai primi due.

Nel frammento (8) invece l'informatore di istruzione alta ricorre ad una serie di passati remoti proprio per conferire alla sua narrazione un maggiore distacco. Anche in questo caso la selezione di un tempo comporta che il parlante tenda ad uniformare tutti i piani della narrazione, ricorrendo sempre all'uso dello stesso tempo anche per le informazioni inserite nella subordinata. In tal senso la distinzione nei livelli pragmatico-narrativi è affidata all'articolazione sintattica del periodo più che all'uso dei tempi verbali:

- (8) il periodo di guerra si era / *si era sfollati. si scappò* da Trapani /// eh:: la mia famiglia /  
**E1** **E2**  
 e tutti mh: *andarono via* da Trapani perché *cominciarono* i bombardamenti, quindi  
**E3** **S1**  
 uno sfollamento totale. / la mia famiglia / *venne* chiaramente qua nella zona di  
**E4**  
 Valderiçe lì nel: / baglio antico di Marotta (N., istr. alta, Custonaci – Trapani)

In altri casi, l'alternanza tra l'uso del passato remoto e del passato prossimo sembra fondarsi piuttosto sulla ripercussione nel presente che hanno gli eventi narrati. Ad esempio in (9), l'informatore opera una distinzione dei piani narrativi, ancora una volta, non tanto in base alla distanza temporale quanto agli effetti che le azioni narrate hanno sul presente, i passati remoti fanno avanzare la storia, i passati



- (11) La mia ragazza ha due anni che stiamo assieme. *L'ho conosciuta* alle scuole  
**E1**  
 elementari, *abbiamo fatto* delle recite assieme pure da piccoli, **poi** alle scuole medie  
**E2**  
*ci siamo divisi*. In terza media per gli esami niente | **poi ci siamo ritrovati** in una stessa  
**E3** **E4**  
 comitiva l'estate del 1997. (F., istr. media, Palermo).  
**MR1**
- (12) e allora io / finita la scuola serale - *ho frequentato*: un anno di scuola: alberghiera - a  
**MR1** **E1**  
 Cefalù di ottoçento ore- [...] *mi son preso* l'attestato [...] - *ho lavorato* per un anno e  
**E2** **E3**  
 mezzo in una rosticceria - dove li *ho avuto* esperienza da pizzaiolo e da cuoco. dopo  
**E4**  
 di quello / *ho iniziato a lavorare* in un centro stampa.  
**E5**  
 (F., istr. media, Termini Imerese - Palermo)

Differentemente, i parlanti giovani di istruzione alta, tendono ad organizzare le sequenze narrative mettendo in evidenza non solo la concatenazione lineare degli eventi ma anche riportando, grazie all'uso di subordinate, gli opportuni riferimenti alla relazione tra le azioni. Dai frammenti (13) e (14), seppur connotati da una difficoltà di pianificazione testuale, che si evince dalle ripetizioni e dai frequenti cambiamenti di progetto sintattico, che conferiscono al testo un andamento più incerto, è possibile osservare come le subordinate temporali oltre a svolgere la funzione di aggancio temporale servono proprio a creare lo sfondo narrativo in cui si situano gli eventi narrati.

In questi ultimi due frammenti esaminati la narrazione viene spesso inframmezzata da espansioni esplicative all'imperfetto che permettono di articolare opportunamente la testualità in più piani e, al tempo stesso, riflettono una modalità di pianificazione del testo, che sembra essere caratteristica degli informatori giovani più che di quelli anziani, per cui ogni evento descritto viene affiancato da una digressione esplicativa. Sembra, dunque, che da una parte i parlanti anziani di istruzione alta procedono con una attenzione per la gerarchia delle azioni che guarda soprattutto alla successione cronologica degli eventi narrati, dall'altra i parlanti giovani alternano in modo più libero informazioni di sfondo e di primo piano, affidando proprio alla distinzione dei livelli pragmatico-narrativi la ricostruzione della loro salienza narrativa:

- (13) sì. allora. eh:: quando:: andavo in terza media eh: *hanno organizzato* uno: uno  
 [ S1 ] E1  
 scambio culturale con una scuola:: eh:: di Grano+ di Granobl. eh:: e praticamente |  
 cioè erano ragazzi: mh: / che erano più piccoli di noi, avevano un anno in mmeno, eh::  
 [ S2 ]  
 eh niente: *hanno organizzato*: questo scambio, eh::: prima *sono venuti* i Francesi da  
 E2 = E1 E3  
 noi, eh:: però: non qui a Sclafani: *sono venuti* a Caltavuturo. eh:: perché siccome noi  
 E4=E3 S3=  
*eravamo* pochi non: non pote+ | cioè non *potevano* ospitarli.  
 =S3  
 (F., istr. alta, Caltavuturo – Palermo)

- (14) tre anni fa un incidente stradale. / mentre: *uscivamo* dalla discoteca eh:: un ragazzo mi  
 MRI [ S1 ]  
*ha dato* un passaggio / per anda+ | per:: ritornare a casa. // il problema è che io *avevo*.  
 E1 [ S2 ]  
 premura perché mia madre non:: non lo *sapeva* che io // *ero* in questo posto. a:: | in  
 [ S3 ] [ S4 ]  
 questa discoteca a bballare. questo ragazzo *era* / ubriaco mentre *scendevamo* per:: |  
 [ S5 ] [ S6 ]  
 cioè *andavamo* a casa e il: | erano le tre le quattro del mattino // a tterra c'era il  
 [ MR2 ] [ S7 ]  
 bagnato, la macchina è *sbandata* e *siamo andati a finire* contro un muro.  
 [ E3 ] E4  
 (FIV., istr. alta, Caltavuturo–Palermo)

In particolare, nel caso di (14) si può osservare come vi sia un sbilanciamento tra tutte le sequenze che costituiscono lo sfondo narrativo e gli eventi, a favore delle prime, dal momento che l'informatrice antepone tutta la cornice con le informazioni di *background* alla sequenza lineare degli eventi narrativi che trova spazio alla fine del frammento riportato.

## 5. Conclusioni

Da quanto sin qui osservato, è possibile avanzare qualche spunto di riflessione che sembra emergere dal corpus esaminato sia in relazione al tipo di articolazione dei piani narrativi, e dunque all'intreccio tra cognizione e lingua, sia più in generale rispetto all'appartenenza dell'italiano tout court alle lingue tipologicamente esocentriche.

Relativamente al primo punto, come abbiamo appena osservato emerge un diverso rilievo nella strutturazione del periodo nei parlanti di istruzione alta a

seconda che appartengano alla generazione dei nonni o a quella dei figli. Infatti, benché entrambi siano accomunati dalla tendenza a creare, principalmente grazie all'uso dell'imperfetto in subordinate, un *backgrounding* in cui fare insistere le informazioni di primo piano, cambia la modalità con cui questo sfondo interagisce con la linea principale degli eventi: più equilibrato e tendenzialmente lineare nei parlanti anziani che sembrano essere più abili narratori riuscendo a pianificare il testo in modo da dosare i riferimenti alla relazione tra gli eventi; più dinamico e intrecciato nei giovani che affidano una parte consistente di informazioni ad espansioni che si intersecano con le informazioni di primo piano.

L'osservazione dei tempi del testo, alla luce dei quadri teorici che abbiamo usato come riferimenti, ha portato dunque a poter identificare dei profili di costruzione testuale che dipendono direttamente dalla variabile diagenazionale. Analogamente, l'attenzione per la variabile diastratica ci ha permesso di osservare come, a prescindere dall'età degli informatori, i parlanti di istruzione bassa tendano ad appiattire alla sola linea degli eventi la narrazione senza concedere spazio al *background*. Come abbiamo osservato, ciò comporta che la scelta di un unico tempo verbale sia mantenuta tendenzialmente per l'espressione di tutte le azioni narrate che vengono giustapposte in modo concatenativo. La padronanza del codice utilizzato sembra intervenire dunque in modo sostanziale anche nell'organizzazione cognitiva che dà luogo alla narrazione degli eventi.

L'osservazione dei differenti profili narrativi, legati al livello di istruzione dei parlanti esaminati, ci permette inoltre di affermare che se l'italiano standard, padroneggiato meglio dagli informatori di istruzione medio-alta, possiede a pieno titolo le caratteristiche sintetizzate nella tabella 1 per le lingue esocentriche, diversamente la lingua prodotta dagli informatori di istruzione bassa sembrerebbe avvicinare l'italiano parlato più alle lingue endocentriche, data la semplificazione delle potenzialità morfologiche e flessive del sistema verbale dell'italiano che avviene in questa varietà di lingua. Ciò porta a concludere che, nello studio di una lingua come l'italiano, prospettiva tipologica e prospettiva variazionistica vanno tenute senz'altro insieme per avere una visione che tenga conto non solo della lingua in quanto tale ma anche degli usi effettivi dei parlanti nelle loro varie realizzazioni.

## Bibliografia

- Benveniste, E. 1971. *Problemi di linguistica generale*. Milano: Il Saggiatore.
- Bertinetto, P.M. 1986. *Tempo, aspetto e azione nel verbo italiano. Il sistema dell'indicativo*. Firenze: Accademia della Crusca.
- Bertinetto, P.M. & Squartini, M. 1996. La distribuzione del Perfetto Semplice e del Perfetto Composto nelle diverse varietà di italiano. *Romance Philology* 49: 384-419.

- Centineo, G. 1991. Tense-Switching in Italian: The Alternation between passato prossimo e passato remoto in Oral Narratives. In S. Fleischman & L.R. Waugh (eds), *Discourse-Pragmatics and the Verb: The Evidence from Romance*. London & New York: Routledge, 55-85.
- Cresti, E. 2000. *Corpus di italiano parlato*, vol. I. Firenze: Accademia della Crusca.
- D'Agostino, M. & Paternostro, G. (eds) 2006. *Costruendo i dati. Metodi di raccolta, revisione e organizzazione della banca dati della sezione sociovariazionale*. Materiali e ricerche dell'Atlante Linguistico della Sicilia, 17. Palermo: Centro di Studi filologici e linguistici siciliani.
- Hopper, P. 1979. Aspect and foregrounding in discourse. In T. Givón (ed.), *Discourse and Syntax*. New York: Academic Press, 213-42.
- Jansen, H., Jensen, B.L., Jensen, E.S., Korzen, I., Polito, P., Skytte, G. & Strudsholm, E. 1997. Testi paralleli scritti e orali, in italiano e danese. Strategie narrative. *Cuadernos de Filologia Italiana* 4: 41-63.
- Korzen, I. 2002. Il Trapassato Prossimo in un'ottica pragmatico-testuale. In H. Jansen, P. Polito, L. Schøsler & E. Strudsholm (eds), *L'infinito & oltre. In omaggio a Gunver Skytte*. Odense: Odense University Press, 203-226.
- Korzen I. 2004. Dalla microstruttura alla macrostruttura. In P. D'Achille (ed.), *Generi, architetture e forme testual*. Firenze: Franco Cesati Editore, 363-376.
- Korzen I. 2005. Struttura linguistica e schema cognitivo: tipologie a confronto. *Copenhagen Studies in Language* 31, I: 123-134.
- Lo Cascio, V. 1982. Temporal Deixis and Anaphor in Sentence and Text: Finding a Reference Time. *Journal of Italian Linguistics* 7: 31-70.
- Lo Cascio, V. & Vet, C. (eds) 1986. *Temporal Structure in Sentence and Discourse*. Dordrecht: Foris Publications.
- Paternostro, G. 2004. Un parlante, tanti parlati. Riflessioni sulle produzioni delle fonti nella geolinguistica variazionista. *Rivista Italiana di Dialettologia. Lingue dialetti società* XXVIII: 45-86.
- Weinrich, H. 1978. *Tempus. Le funzioni dei tempi nel testo*. Bologna: Il Mulino.



# DA UNA LINGUA ENDOCENTRICA AD UNA LINGUA ESOCENTRICA? IL CASO DELL'ITALIANO

Emanuela Cresti

LABLITA, Università di Firenze

## 1. Premessa

Il presente contributo offre una veloce sintesi dei principali caratteri strutturali dell'italiano, risultato di un lungo percorso di separazione dal latino, al fine di sollevare la questione di come sia possibile la derivazione di una lingua esocentrica, come appare l'italiano insieme alla gran parte delle altre lingue romanze, da una lingua endocentrica, come viene classificato il latino. Nostro intento non è quello di offrire una soluzione o un'ipotesi a questo proposito, ma molto modestamente quello di presentare un quadro entro il quale il problema possa meglio emergere e rendersi evidente.

Le mie ricerche ormai più che trentennali sono state dedicate allo studio della lingua parlata (in maniera specifica all'italiano, ma anche alle altre lingue romanze), con ricerche *corpus-based* e sperimentali, vertenti in particolare sui rapporti tra prosodia e struttura informativa (Cresti 2000). Tuttavia fare ricerca sull'italiano parlato implica uno studio della variazione diamesica, con una riflessione sull'impatto evidente che l'uso orale novecentesco ha avuto sulla lingua scritta e sul sistema italiano tout-court, e con la constatazione di quanto sia stata veloce la variazione diacronica connessa. L'insieme di tali osservazioni mi ha portato più in generale a pormi il quesito della formazione e dell'identità dell'italiano. Sono arrivata all'ipotesi semplice, per certi versi banale, ma che ugualmente non trova tutti d'accordo: la formazione di un nuovo sistema linguistico da uno già esistente può essere la più varia e complessa per quanto concerne le vicende storiche e sociali determinanti, ma non può saltare il punto nodale di un uso continuato e totale, soprattutto parlato, del sistema di partenza<sup>1</sup>. Sinteticamente: un nuovo sistema nasce dalla "parole" di uno vecchio.

---

<sup>1</sup> Può essere esemplare la vicenda dell'ebraico che ormai è la lingua materna di una larga comunità di parlanti, ma che si è trasformata da lingua classica della Torah, da lingua di Cresti E., Korzen I. (eds), *Language, Cognition and Identity. Extensions of the endocentric / exocentric language typology* © 2010 Firenze University Press.

## 2. Il sistema italiano

L'italiano ha nascita tarda, possiamo indicare la metà del XIII secolo, come anche la recente e imponente "Grammatica dell'italiano antico" (GIA) di Salvi e Renzi ribadisce. In essa i curatori prendono il 1260-61, data di composizione della "Rettorica" di Brunetto Latini, come l'inizio documentale del fiorentino e quindi, per le note vicende, dell'italiano. A questo proposito viene ricordato che se già agli inizi del secolo è presente una serie di testi in fiorentino dopo la metà dello stesso "il flusso si fa ininterrotto"<sup>2</sup>.

Al di là delle vicende linguistiche fiorentine possiamo convenzionalmente prendere l'inizio del XIII secolo come data a partire dalla quale si considerano esistenti un gruppo di volgari italiani, interpretabili, all'interno di una prospettiva storica rovesciata partente dall'oggi, come varietà diatopiche di un sistema in via di formazione, unificazione, e standardizzazione, ciascuno dei quali appare già corredato di interne variazioni diafasiche, tra le quali acquista rilievo la presenza di un uso pubblico scritto (transazioni mercantili, inventari, conti, atti notarili, atti di processi, versioni volgari di statuti comunali), ma con il finire del secolo anche di registri alti di uso letterario<sup>3</sup>. Naturalmente la derivazione dalla lingua madre non è stata un travaso uniforme, ma al contrario è passata attraverso fenomeni di frammentazione e centrifugazione massima dopo lo sfaldamento dell'impero romano di Occidente, per arrivare lentamente a nuove forme di concentrazione e standardizzazione su basi culturali, religiose, sociali, economiche diverse, dando luogo appunto ad un gruppo di Volgari illustri tra i quali è emerso quello fiorentino<sup>4</sup>. Il distacco dal latino e la conseguente formazione dell'insieme di volgari italiani, è stato però un processo lungo, più lungo di quello delle altre varietà dell'universo romanzo, perché ha richiesto quasi un millennio per giungere a compimento, anche se la fisionomia dei diversi volgari può dirsi profilata già dal X secolo<sup>5</sup>.

---

cultura e di tipo veicolare per fini commerciali nelle regioni della Palestina, in una lingua viva solo dopo che, anche se in maniera artificiale e volontaria, ma molto prima della costituzione dello stato di Israele, è stata impiegata come lingua parlata di uso quotidiano. Si veda Izre'el & Mendelson (2002).

<sup>2</sup> Salvi & Renzi, op. cit., p. 9

<sup>3</sup> Vorrei ricordare a questo proposito il recente Storia linguistica di Venezia (2010) di Tomasini, che ben documenta il tardo uso letterario del Volgare di Venezia, dopo quello burocratico e mercantile precedentemente consolidato.

<sup>4</sup> Per tutti i riferimenti relativi alla storia della lingua italiana e alla formazione del suo sistema linguistico si vedano Migliorini (1961) e Rohlfs (1966-69).

<sup>5</sup> Tuttavia non bisogna dimenticare che la vicenda di una coscienza chiara e provata della nuova identità linguistica e dell'avvenuta separazione dal sistema "latino" non può dirsi acquisita neppure con la composizione della geniale "Grammaticetta" dell'Alberti del 1437-38, che rimane un *unicum* e giace manoscritta per cinque secoli. Solo le conferme grammaticali del primo '500, per opera di Bembo e di Fortunio, e poi tutta la riflessione

La prima caratteristica da evidenziare in questo processo è una derivazione senza soluzione di continuità del sistema fiorentino e poi italiano da quello latino, che pure è la comune lingua madre romanza, ma il carattere di filiazione diretta e totale costituisce una specificità non condivisa nella stessa misura da nessuna delle altre lingue del gruppo<sup>6</sup>. A questo proposito però, forse è necessaria una cautela che ci deriva dall'avvertimento in apertura della GIA, nella quale si fa notare che il preteso isomorfismo dei sistemi "fiorentino antico" e "italiano contemporaneo" è un'illusione, perché le variazioni intervenute sono tante e riguardano tutti i sottosistemi (fonologico, morfologico, sintattico, semantico). Quindi anche se la variazione dell'italiano è stata compressa e controllata per diversi secoli, il cambiamento strutturale comunque si è verificato, e potrebbe darsi quindi che il sistema dell'italiano contemporaneo non corrispondesse più ad una discendenza ininterrotta dal latino. Tuttavia crediamo che quelle diversità dell'attuale sistema, evidenziate giustamente nella GIA, siano emerse o si siano intensificate soprattutto nelle ultime decadi, a nostro avviso proprio sull'onda di quell'uso totale e primariamente parlato, che come dicevamo, è di per sé portatore di mutazione. D'altro canto si può constatare che tali mutamenti non arrivano ad inficiare la linea diretta di derivazione. Proprio i dati presentati dalle ricerche della GIA confermano da un lato che già alla metà del XIII secolo nel volgare fiorentino lo stacco dal sistema latino era compiuto e d'altro canto che alcuni aspetti fondamentali di tale

---

grammaticale rinascimentale sono necessarie a sancire il riconoscimento pieno dell'identità linguistica dell'italiano.

<sup>6</sup> A questo proposito forse val la pena di ricordare che ciò che corrisponde alla denominazione di "lingue romanze" non si riduce alle 5 lingue ufficiali di uno stato (italiano, francese, spagnolo, portoghese, rumeno) ma ad una quarantina tra gruppi e sottogruppi linguistici, con almeno 800 milioni di parlanti. Devono essere annoverate lingue che sono considerate insieme ad altre ufficiali in un certo stato, come il catalano per la Spagna, o non sono lingue ufficiali ma aspirano allo statuto di lingue, come le lingue sarde e ladine in Italia, o i gruppi dialettali che non emergono come lingue ma che hanno chiare caratteristiche idiosincratiche e raggruppano un insieme di parlate come i dialetti gallo-italici, quelli franco-provenzali, quelli campani, quelli siciliani, quelli dell'Estremadura, ecc.. E' bene però considerare in questo elenco anche lingue ormai estinte, che sono state però importanti veicoli di comunicazione, come la lingua franca parlata in tutti i porti del Mediterraneo tra il XII e ancora il XIX secolo, o il Mozarabo facente capo a vari dialetti romanzi parlati nella penisola iberica all'epoca della Reconquista, influenzati dall'arabo e scritti in caratteri arabi, o le lingue giudeo romanze parlate dalle varie comunità ebraiche in diversi centri e regioni italiani e spagnoli a partire dal X secolo. Se poi valichiamo l'oceano oltre alle lingue ufficiali con le loro varietà, riconosciute nei vari stati latino americani, come lo spagnolo e il portoghese, possiamo citare il Cocoliche parlato dagli italiani di Buenos Aires, il Talians parlato dalle comunità venete nel sud del Brasile. In un capitolo a parte vanno poi segnalate le cosiddette lingue creole, nate da contatti recenti tra una lingua di colonizzazione e le parlate indigene o di comunità di schiavi importati, come il creolo haitiano su base francese, il creolo capoverdino su base portoghese, ma anche lingue come il Calò derivato dal contatto tra lo spagnolo, componenti arabe e la lingua dei Rom stanziati in Andalusia.

nuovo sistema sono rimasti costanti in quello che è l'esito linguistico attuale. Vorrei cercare di evidenziare quindi alcuni "super caratteri" che sembrano comuni sia al fiorentino antico che all'italiano odierno e che in blocco derivano ma al tempo stesso si oppongono al sistema latino, il quale a sua volta viene preso a riferimento come un sistema monolitico e sincronico, il che è evidentemente solo un'ipotesi di comodo, che deve essere riconsiderata.

Cerchiamo di evidenziare quindi alcuni punti di distacco strutturale molto evidenti, e che già a metà del '200 testimoniano della raggiunta separazione; ne indichiamo come causa primaria l'ipotesi elementare che sia stato l'uso orale del latino in diafasia pragmatica e bassa diastratia, oltre all'uso fattone da parlanti di madrelingua diversa, a portare ad un generale processo che chiameremo di "creolizzazione lata".

Il primo sistema a rimanerne coinvolto appare la materia linguistica di base, ovvero i suoni, che appaiono fortemente differenziati da quelli del latino<sup>7</sup>. Va detto peraltro che il sistema vocalico italiano costituisce nel suo complesso un supporto melodico idiosincratico, chiaramente distinto anche da quelli delle altre sorelle romanze. Esso annovera infatti il passaggio da un sistema vocalico a tre gradi di apertura e con differenze di lunghezza (per 10 vocali), indipendenti dalla tonicità, ad un sistema a quattro gradi apertura e differenze timbriche (per 7 vocali) dipendenti dalla tonicità, con una variante non tonica a 3 gradi (per 5 vocali). Ad esso si aggiunge il passaggio da un sistema di pochi dittonghi pienamente vocalici del latino (*ae, au, oe*) presto perduti a due serie di dittonghi, tra le quali la serie ascendente (composta da una consonante approssimante *j* o *w* e da una vocale), assente in latino, è la serie principale e produttiva dell'italiano, e da una serie discendente vocalica di minor frequenza e con restrizioni distributive. La trasformazione dell'intero sistema vocalico è strettamente connessa alla impossibilità di distinzione casuale per la flessione nominale. Ma anche i caratteri consonantici, ovvero i "rumori" che accompagnano le vocali per la formazione delle sillabe, sono diversi: la caratteristica fondamentale sembra quella legata ad un arricchimento dei luoghi articolatori (le postalveolari e/o palatali), ma anche delle modalità (le affricate e le approssimanti), con un ampliamento nel numero delle fricative, una diffusa duplicazione dei fonemi nelle loro varianti sorde e sonore, all'interno di un generale processo di sonorizzazione per influenza germanica, nonché il sistematico impiego dell'intensità, in parte derivata da processi di assimilazione di gruppi consonantici latini. Devono essere aggiunti poi i diversi valori dell'accento, ma soprattutto i diversi criteri di composizione sillabica e di forma di parola, che portano alla perdita delle desinenze consonantiche sia nominali che verbali.

La ristrutturazione fonologica e la nuova forma di parola sono certo una componente entro il complesso sistema di influenze reciproche che hanno come

---

<sup>7</sup> Si veda Tagliavini (1949).

esito morfologico la perdita della declinazione casuale dei nomi e la ristrutturazione del sistema verbale, effetti che tuttavia si inquadrano all'interno del processo proprio delle creolizzazioni che producono la perdita o una drastica riduzione della morfologia flessiva della lingua base. Ma mentre nelle lingue creole la lingua base è ridotta anche nella sue funzioni sociali e culturali e quindi nel suo lessico, questo è avvenuto solo in parte per il latino, che ha mantenuto fondamentali e rilevanti usi scritti e colti, anche se confinati ad una percentuale molto ristretta di utenti. Il cambiamento sembra essersi conformato in un passaggio dalla lingua sintetica di base a tante lingue analitiche, che hanno sostituito le funzioni della declinazione e dei morfemi flessivi con morfemi liberi (articoli, pronomi, preposizioni, congiunzioni, clitici, forme verbali ausiliarie). Si veda la declinazione casuale sostituita da sintagmi preposizionali e per la flessione verbale, che si è tuttavia mantenuta consistente nei tempi semplici, la formazione di tempi composti, di diatesi, di aspetti azionali e perifrastici tramite verbi ausiliari. Rilevanti tra i morfemi liberi il sistema degli articoli determinativi e indeterminativi, e i clitici in tutta la loro gamma funzionale (anaforica, lessicale, flessiva verbale)<sup>8</sup>, entrambi assenti in latino. Inoltre anche la formazione di parola, che compete alla morfologia lessicale, passa da una formazione esclusivamente derivazionale (suffissi, prefissi, prefissoidi) del latino, che si differenzia per questo carattere in maniera così netta dal greco, ad una sempre più diffusa formazione per composti<sup>9</sup>.

Le due questioni fondamentali concernenti la sintassi: l'ordine delle parole entro la frase e il sistema della subordinazione, presentano anch'esse forti mutazioni. La fine del sistema flessivo casuale ha tuttavia effetti a cascata e rimane una delle cause determinanti anche nel dominio sintattico. In primo luogo la perdita flessiva dei nomi ha prodotto una restrizione distributiva dell'italiano, con la necessità di marcare il fondamentale ruolo di oggetto attraverso l'ordine VO, per i gruppi nominali saturanti la valenza verbale. Condizione che appare la novità rilevante rispetto ad un ordine OV, ma in ultima analisi libero del latino. Per quanto riguarda la subordinazione i fenomeni sono numerosi ma le forme di subordinazione più diffuse (oggettive, soggettive) sono passate da forme nominalizzate del latino (accusativo + infinito), anch'esse connesse alla segnalazione casuale, a forme di subordinazione esplicita introdotta da un morfema libero, come il complementatore polivalente 'che' in italiano. Se si pensa poi che quasi il 40% della subordinazione italiana è costituito da frasi relative esplicite, con forme verbali di modo finito e ancora introdotte da *che*, il confronto con il latino va nel senso del molto più basso rilievo quantitativo delle relative in generale e delle esplicite in particolare, le quali

---

<sup>8</sup> Si veda Cresti (2009).

<sup>9</sup> Tuttavia la formazione di parola per composizione cresce con il procedere dei secoli e diviene percentualmente rilevante dopo il '700 per influenza prima del francese e poi del tedesco e dell'inglese, ma anche del linguaggio scientifico e tecnologico che porta alla coniazione di composti neo-classici. Si veda per il rapporto con il francese Dardi (1992).

ultime vengono di preferenza restituite tramite nominalizzazione, con participi presenti e gerundi, ancora forme accordate per caso<sup>10</sup>. Bisogna aggiungere poi a questo quadro il complesso di quei costrutti che vengono indicati come sintassi segmentata o della focalizzazione<sup>11</sup> (scisse: ‘è in vacanza che voglio andare’, ‘non è che mi piaccia molto’, ‘è perché va bene così’; pseudo-scisse: ‘quello che è importante è che lui ce la faccia’; dislocate a destra e sinistra: ‘lo voglio bello forte il caffè’, ‘il giornale l’ho già comprato’), che sembrano anch’essi una risultante dell’uso parlato, con il rilievo dato a strutture informative. Tuttavia costrutti così conformati appaiono realizzabili solo in un sistema per un verso fornito di quella morfologia libera (clitici, ausiliari, complementatori polivalenti) che abbiamo visto mancava in latino e dall’altro di un sistema sintattico fondato sulla subordinazione esplicita.

Le veloci osservazioni relative al sistema fonetico, morfologico e sintattico dell’italiano, tutte presenti già nel fiorentino antico, appaiono sufficienti a provare l’avvenuto distacco dal latino e la realizzazione di un sistema lingua diverso fin dalla metà del ’200. Rimane tuttavia non affrontato il sistema lessicale che però è il settore che meglio può evidenziare, anche per un non esperto, l’identità di una lingua. Il lessico, infatti, in maniera complessiva cattura la realtà sia come specchio di una concezione del mondo condivisa da una comunità di parlanti, sia come filtro che vaglia ciò che è importante e ciò che è effimero, consegnando alla storia la vicenda di una lingua.

### 3. La derivazione latina del lessico italiano

Anche il lessico, come gli altri sottosistemi dell’italiano, deriva in maniera diretta dal latino. Gli studi che testimoniano la discendenza sono numerosissimi a cominciare dall’opera fondamentale di Migliorini<sup>12</sup> per arrivare al Lessico Etimologico Italiano (LEI), il dizionario etimologico della lingua italiana che permetterà di collocare il lessico italiano in un contesto romanzo complessivo. L’impresa ideata e coordinata da Max Pfister, è tuttavia ancora lontana dalla sua conclusione<sup>13</sup>. È nostro intento tuttavia affrontare la questione da una prospettiva che ci viene fornita dalla linguistica dei corpora.

Sappiamo che in lingue di cultura con lunga tradizione scritta il lessico può superare le 300 mila entrate, perché i lemmi si estendono a comprendere da un lato termini arcaici, magari ormai fuori uso ma presenti nella tradizione testuale,

<sup>10</sup> Si veda Scarano (2003).

<sup>11</sup> Si veda Panunzi (2010).

<sup>12</sup> Migliorini. op. cit.

<sup>13</sup> E’ stato da poco stato pubblicato il centesimo fascicolo del LEI, che è arrivato alla lettera D. La conclusione è prevista per il 2032.

dall'altro le terminologie tecniche che continuamente si ampliano e rinnovano e i prestiti. Ma entro questo *mare magnum* è necessario procedere attraverso una distinzione tra i lemmi semantici veri e propri e quei lemmi funzionali, i morfemi liberi, che non sono parole semantiche piene. Come abbiamo anticipato i morfemi liberi dell'italiano sono nuovi rispetto al latino o comunque fortemente differenziati nella loro forma, funzione e valore semantico, ma il loro apporto può essere meglio valutato entro la morfologia e la sintassi piuttosto che nel lessico, visto che oltre tutto essi assommano a poche centinaia appartenendo a classi chiuse<sup>14</sup>. Invece per quanto riguarda il "grosso" del lessico restante bisogna procedere introducendo la fondamentale distinzione grammaticale e semantica, ma in ultima istanza cognitiva, tra ciò che è verbo e ciò che non lo è, e che corrisponde alle classi lessicali in primis dei nomi e poi degli aggettivi e degli avverbi. Se i nomi, includendo in essi i nomi propri, possono annoverare entro il dizionario di una lingua centinaia di migliaia di lemmi, e diverse migliaia ne contano gli aggettivi e gli avverbi, solo alcune migliaia di lemmi sono verbali.

I dati quantitativi possono essere desunti dalla composizione lessicale di un corpus che è preso in genere a esempio di rappresentatività di una lingua nazionale: il *British National Corpus* (BNC). Esso corrisponde a 100 milioni di *tokens*, che proiettano un lessico di circa 80.000 lemmi (nostre stime sulle entrate con più di 5 occorrenze). Evidentemente i corpora non sono dizionari che tendono all'esaustività, presentando tutti i lemmi di un sistema lingua, ma tendono ad essere rappresentativi dell'uso, dal quale per forza di cose rimangono per lo più escluse le punte estreme dell'arcaicità e della tecnologia. Ebbene il lessico verbale del BNC rappresentato nella nostra stima annovera un repertorio di circa 7000 verbi.

Per quanto concerne la derivazione del lessico italiano da quello latino è noto che essa può essere considerata diretta ma non avviene tanto e solo dal lessico classico, quanto in una parte significativa da un lessico latino tardo. Ed è molto ben documentato fin dai primi anni dell'era cristiana il significativo apporto di lingue diverse: imponenti prestiti dalle lingue germaniche (goti, longobardi, franchi), prestiti greci ellenistici e bizantini, ebraici, arabi. Ma di particolare rilievo appare il contributo di ciò che può essere chiamato nel suo complesso il "cristianesimo", perché anche se non è corretto identificare il lessico latino tardo con un lessico cristiano *tout-court*, la teoresi e la prassi religiosa sono in ogni caso da valutare come tra le principali componenti di rinnovamento. Esse comportano insieme: varietà diafasica e diamesica legata alla predicazione e alla preghiera, ma anche varietà diastratica di registro basso, quello proprio della gran massa dei credenti, anche se nel IV secolo la gran parte della classe senatoria si era convertita al

---

<sup>14</sup> Naturalmente deve essere ben distinta la formazione di un lessico o dizionario di una lingua e le occorrenze dei lemmi in un corpus. Infatti la frequenza di occorrenza dei morfemi liberi è altissima e può arrivare a coprire l'80% delle occorrenze di un corpus.

cristianesimo. Ma non bisogna dimenticare la varietà diatopica del latino coniato e parlato in Africa settentrionale, dove si era per prima e largamente diffusa la religione cristiana, e la varietà di contatto per le traduzioni dal greco giudaico dei primi Vangeli e di tutta la produzione apocrifia, tradotti in un latino sempre di varietà africana. Del resto i massimi esponenti antichi del pensiero cristiano (Origene, Tertulliano, Agostino) sono originari dell’Africa, e lo erano anche i più grandi grammatici della tardo ellenismo (Apollonio, Donato, Prisciano) ed infine Girolamo, il traduttore e compositore della Vulgata, che divenne il testo di riferimento del latino tardo e dell’alto medioevo, non era di origine romana ed era stato allievo di Donato.

Arriviamo così alle soglie di quel X secolo per il quale è possibile ipotizzare la prima gestazione dei diversi Volgari italiani, con un lessico latino la cui composizione è ormai lontana da quella del latino classico. Tuttavia se si scorrono le liste dei prestiti o delle neo-formazioni latine di registro basso, agricolo e domestico, o di ambito religioso, si noterà che il numero dei verbi di nuova base semantica sono pochi (‘manducare’, ‘nevicare’, ‘pettinare’). I verbi “cristiani” sono legati per lo più alla liturgia (‘battezzare’, ‘benedire’) e non sembrano costituire un repertorio troppo significativo. Si pensi poi che entro i prestiti arabi che importano in maniera duratura centinaia di lemmi in tutti campi semantici, forse l’unico verbo riconosciuto è ‘ricamare’, che come si capisce non è proprio determinante e appare anche in arabo un denominale (da tessuto). Forse più rilevante il contributo longobardo (‘graffiare’, ‘russare’, ‘scherzare’, ‘spaccare’, ‘tuffare’). In conclusione se il lessico nominale che arriva in dote ai volgari italiani è in buona parte diverso da quello del latino classico, il lessico verbale rimane in sostanza un lessico latino.

#### 4. Il lessico verbale

All’interno del lessico verbale, così contenuto numericamente e caratterizzato come abbiamo anticipato in maniera tendenzialmente conservativa, bisogna in ogni caso procedere per identificare classi semantiche che hanno funzioni e rilievo molto diverso nella produzione linguistica. Per esempio devono essere considerati a parte quei verbi veri e propri ausiliari (‘essere’, ‘avere’), ma anche gli usi di ‘essere’ di tipo predicativo copulare (attributivo, identificativo, locativo)<sup>15</sup>, i verbi modali, ma anche i verbi a reggenza frastica (*dicendi, putandi, sentiendi*). Per quanto riguarda i verbi “ausiliari” essi sono tutti di derivazione latina: ‘essere’, ‘avere’, ‘potere’, ‘dovere’, ‘volere’. Ma anche i più importanti verbi a reggenza frastica: ‘dico’, ‘vedo’, ‘sento’, ‘credo’, ‘spero’, ‘prometto’ sono verbi latini, indipendentemente dal fatto che la subordinazione da essi introdotta abbia in italiano struttura esplicita e

---

<sup>15</sup> Si veda Panunzi (2010).



non più nominale come in latino. Le forme di queste classi verbali hanno frequenza molto alta perché, avendo funzioni ausiliarie, servono di appoggio temporale, aspettuale, modale, di mediazione evidenziale di altri contenuti eventivi di cui si predica. Esse non sono però pienamente rappresentative dei campi semantici eventivi di una lingua e del modo di concepire un evento.

Il gruppo dei verbi pienamente semantici si riduce quindi ad un macro-gruppo di verbi azionali, all'interno dei quali esiste un'ulteriore e chiara distinzione tra verbi *di moto* e verbi *non di moto* o più propriamente di "azione". Le ricerche su corpora mostrano che in effetti i verbi azionali nel loro complesso, di moto e non, dividono in maniera quasi paritaria le occorrenze con i verbi ausiliari; al loro interno, poi, i verbi di moto e non, si spartiscono altrettanto paritariamente le frequenze. Appare quindi significativa anche per il suo rilievo quantitativo, la classificazione delle lingue proposta da Talmy sulla base di ciò che egli chiama "lessicalizzazione" verbale e che dipende da fondamentali tratti semantici dei verbi di moto<sup>16</sup>.

L'ipotesi di Talmy è che l'espressione del movimento può essere regolata da tratti di direzione, come nel caso italiano 'andare', che implicano sempre 'da x a y', oppure da tratti di maniera del movimento, come 'strisciare', che non implicano la direzione. La classificazione prende spunto dalla denominazione tipologica proposta da Bloomfield (1933), sulla base di caratteristiche morfo-sintattiche, di lingue endocentriche ed esocentriche, venendo estesa alla considerazione dei peculiari aspetti semantici dei verbi di moto.

Dato che l'espressione della direzione rimane tuttavia un aspetto semantico ineliminabile, le lingue che hanno verbi di moto con maniera esprimono la direzione attraverso un satellite, ovvero una preposizione/avverbio che di preferenza segue il verbo e forma con esso un predicato complesso, che porta ad un uso inaccusativo e telico del verbo stesso. Herslund (2005), Baron (2005), e Korzen (2005a; 2005b) hanno dedicato importanti studi comparativi tra le lingue germaniche, in particolare il danese, considerate nel loro complesso lingue endocentriche, e le lingue romanze, considerate lingue esocentriche, con studi specifici volta a volta sul francese, lo spagnolo e l'italiano<sup>17</sup>.

La classificazione in lingue endocentriche e esocentriche, fondata sulle caratteristiche dei verbi di moto, permette in realtà di evidenziare una serie di ulteriori correlazioni lessicali, ma anche morfo-sintattiche, che sono proprie delle lingue interessate. La classificazione del lessico di movimento sembra avere conseguenze o comunque correlazioni anche con il lessico nominale, che può risultare implicato nei ruoli argomentali di tali verbi, secondo una strategia

---

<sup>16</sup> Si veda Talmy (1985 e 2000).

<sup>17</sup> Alla denominazione di lingue endocentriche ed esocentriche proposta dalla scuola danese, la letteratura affianca proprio per la caratteristica dell'espressione della direzione attraverso satelliti la dizione lingue *Satellite-Framed* e *Verb-Framed* (Talmy 2000), oppure più recentemente *Path-in-Verb* e *Path-in-Non-Verb* (Slobin 2004).

complementare: a prevalenza di verbi più astratti, come quelli di moto con direzione, si affiancherebbe una correlazione di nomi specifici o iponimi, a prevalenza di verbi più concreti e specifici, come i verbi di moto con maniera, si affiancherebbe una correlazione di nomi iperonimi, o con incorporazione. Gli studi di Korzen hanno evidenziato che per esempio in danese siamo in presenza di un lessico nominale con un maggior numero di termini iperonimi (*bil* ‘mezzo di trasporto’) con incorporazione (*person bil, lastbil, varebil, rutebil*), rispetto all’italiano, che avrebbe invece iponimi (‘automobile’, ‘camion’, ‘furgone’, ‘autobus’). Tralascio le importanti implicazioni relative al diverso uso degli articoli, determinativi, indeterminativi e grado zero, nelle due lingue proprio in conseguenza della diversa semantica di nomi e verbi. Korzen (2005a; 2005b) ha anche studiato una serie di caratteristiche e costrutti morfo-sintattici che correlano con la distinzione semantica delle due classi verbali e nominali, evidenziando la maggiore ricchezza flessiva del sistema verbale italiano e la maggiore strutturazione sintattica del periodo italiano con la grande presenza di forme di subordinazione infinitiva e vere e proprie forme di nominalizzazione (sistema esocentrico), che invece sono molto più ridotte per varietà e quantità di uso nel danese (sistema endocentrico).

A nostro avviso deve però essere ampliato il novero degli aspetti che consentono la classificazione in lingue endocentriche ed esocentriche. Come abbiamo visto all’interno del lessico verbale le frequenze sono spartite tra verbi di moto e verbi azionali, i quali ultimi non avevano per il momento ottenuto particolari indagini di tipo sistematico. Devono essere invece prese in considerazione le importanti ricerche condotte da Moneglia sul lessico verbale (Moneglia 2010; Panunzi & Moneglia 2004; Moneglia & Panunzi in questo volume), propriamente azionale, dell’italiano, che sono state confortate anche da dati diacronici per l’italiano, e da alcuni confronti con altre lingue romanze e non.

Lo studioso ha mostrato che all’interno di questo dominio semantico poteva essere concepita una distinzione basata su tratti semantici di azione astratta, per verbi detti da Moneglia verbi generali (‘prendere’, ‘mettere’, ‘dare’, ‘aprire’, ‘attaccare’, ‘girare’, ‘tirare’, ‘alzare’), e tratti semantici inglobanti l’oggetto, o lo strumento, o l’agente (‘pitturare’, ‘pettinare’, ‘stirare’, ‘mangiare’, ‘colare’), o comunque denotanti azioni molto specifiche (‘picchiare’), per verbi detti di “attività”. I verbi azionali generali sono caratterizzati da “variazione primaria”, ovvero ciascuno di essi può denotare tipologie azionali ben differenziate cognitivamente (‘aprire un ombrello’, ‘aprire gli occhi’; ‘prendere un libro dallo scaffale’, ‘prendere il portafoglio (a qualcuno)’; ‘attaccare un bottone’, ‘attaccare un cappotto all’attaccapanni’). Le ricerche su corpora mostrano che essi hanno frequenza d’uso alta, mentre i verbi di attività possono presentare numerosi lemmi ma la loro frequenza è bassa. In conclusione alla luce di tali risultati nella classificazione delle lingue fondata sul lessico verbale la distinzione tra quelle esocentriche ed endocentriche dovrebbe essere estesa anche alla considerazione dei

verbi azionali. Alle lingue romanze, italiano in testa, come lingue esocentriche dovrebbero far fronte le lingue germaniche, come lingue endocentriche, fornite di verbi di moto con maniera e di un alto numero di verbi azionali “specifici” o di “attività” con frequenze non molto alte. Manchiamo di studi sistematici per quanto riguarda i verbi azionali delle lingue germaniche, ma i “carotaggi” fin qui condotti per esempio per il tedesco e il danese sembrano confermare l’ipotesi.

Come abbiamo già visto la presenza in italiano di un lessico verbale di moto con direzione e, noi aggiungiamo, azionale con prevalenza di verbi generali, giustifica una classificazione entro la tipologia esocentrica, per cui il lessico verbale italiano, ma possiamo dire del lessico nel suo complesso data la significatività dei verbi<sup>18</sup>, registrerebbe un cambiamento strutturale rispetto al lessico latino considerato lingua endocentrica. Quindi l’evoluzione semantica del lessico italiano non consisterebbe tanto nella perdita o aggiunta di migliaia di lemmi, di mutamenti o slittamenti semantici entro lemmi derivati direttamente dal latino, ma in un cambiamento nella concezione dell’espressione dell’azione, fenomeno che trascina conseguenze di valenza tipologica per una lingua. La mutata concezione azionale dovrebbe andare nella stessa direzione manifestata dagli altri cambiamenti di sistema dell’italiano rispetto al latino, ovvero quella di una pretesa semplificazione da creolizzazione lata: pochi verbi azionali generali con variazione primaria e verbi di moto con direzione. Quindi a parte chiedersi le ragioni, i tempi e i modi di una trasformazione così peculiare e insieme condivisa dalle tutte le lingue romanze, non sembrerebbero sussistere incertezze al riconoscimento di tale processo.

Devono però essere notati alcuni aspetti sia per quanto riguarda il latino che l’italiano e sollevate alcune questioni, che poniamo agli esperti del settore e alle quali noi non siamo in grado di rispondere. Prima di tutto ci parrebbe importante verificare se in altri casi di “creolizzazione lata”, come abbiamo definito il passaggio dal latino alle lingue romanze, avvengano cambiamenti semantici comparabili che portano alla ristrutturazione del lessico verbale.

Ora, sulla base della sola considerazione dei verbi di moto, Talmy classifica il latino tra le lingue caratterizzate dal tipo “movimento con maniera”. Questa classificazione non sembra leggere in maniera completamente aderente la struttura del lessico verbale latino, che a questo punto deve essere estesa anche alla considerazione dei verbi azionali. Possiamo dire che in latino c’è una predominanza di verbi azionali specifici e di attività ristretta rispetto a verbi di azione generali con forte variazione primaria? A noi sembra che il repertorio dei verbi azionali in latino sia caratterizzabile in termini di generalità e che il loro uso sia frequente e implichi

---

<sup>18</sup> Si deve considerare infatti oltre alla maggiore frequenza di occorrenza di un verbo rispetto ad un nome, l’importanza informativa del verbo, semanticamente dedicato all’espressione degli eventi, che devono essere saturati da attanti e temi individuali, espressi da nomi ad esso subordinati, e la sua centralità costruttiva che in genere domina la configurazione sintattica di frase.

variazione primaria, si pensi a verbi come *agĕre, ferre, mittĕre, dāre, ducĕre, capĕre*. Ma non avendo dati poniamo la questione agli esperti e notiamo che sembrerebbe necessaria una ricerca in tale direzione, con risultati quantitativi desunti da indagini sistematiche su corpora di latino.

D'altro canto è noto che in latino ai verbi di moto con maniera si affiancano verbi di moto con prefissazione di direzione, i cosiddetti preverbi. Il fenomeno rientra nel caso più generale dei *verba composita* del latino che attraverso un preverbo (preposizione/avverbio spaziale e/o temporale prefissato) mutano la natura azionale espressa, per esempio *facio* è atelico e *perficio* è telico<sup>19</sup>. In latino i *verba composita* con preverbo sono molto comuni e il significato che essi veicolano è complessivamente diverso da quello dei verbi semplici di base. In particolare i verbi di moto prefissati non si distinguono solo per un'aggiunta spaziale o direzionale, ma acquistano significati olistici peculiari senza perdere la direzionalità (*ad-eo* 'andare e venire verso', ma anche 'assalire', *sub-eo* 'andar e venire sotto', ma anche 'subire' e 'affrontare'). Siccome l'impiego dei verbi latini di moto prefissati è antico, diffuso e parallelo a quello dei verbi di moto con maniera, non ci pare che questi possano essere equiparati a verbi di moto con maniera accompagnati da satellite (funzione che sarebbe svolta dal preverbo), ma piuttosto a veri e propri verbi di moto con direzione.

In particolare poi deve essere notato che per questi verbi è presente fino da epoche arcaiche, e si mantiene nel latino classico, la possibilità di occorrenza non contigua della preposizione/avverbio con il verbo modificato (*ob vos sacro = obsecro vos*). Anche per i verbi di moto possiamo notare questa possibilità, in Cesare troviamo sia *castris egressi*, dove il verbo prefissato ha reggenza diretta del nome con il caso appropriato di provenienza, sia *e castris Helvetiorum egressi*<sup>20</sup>, dove alla non contiguità si aggiunge però anche la reduplicazione della preposizione *e* che in questo caso è quella che regge il nome con caso ablativo. Rispetto a Virgilio che usa *Tyriam urbem adveni*, col semplice accusativo, in epoca più tarda, per esempio in Girolamo, comincia ad essere prodotto di preferenza, e diventa predominante in Tommaso, un costrutto come *advenio ad Romam*, con reduplicazione della preposizione. Essa in qualche modo appare più equiparabile ad un satellite, perché la caduta della consonante flessiva e la non distinzione della lunghezza vocalica, che nella produzione orale doveva portare a *advenio Roma*, non permetteva più di recuperare il valore funzionale del nome e il verbo di moto con direzione non bastava da solo ad esprimerlo. Quindi solo nel latino tardo, probabilmente, sia i verbi di moto con maniera che quelli di moto con direzione si sono uniformati richiedendo in modo obbligatorio satelliti per esprimere

<sup>19</sup> Rimando all'articolo di Meini & McGillivray (in stampa) che tratta in maniera approfondita e con riscontri su corpora la questione dei *verba composita* e in particolare della frequenza dell'occorrenza contigua o non con il verbo modificato del suo preverbo

<sup>20</sup> Questi esempi sono stati ripresi dall'articolo citato di Meini e McGillivray

compiutamente la direzione. In conclusione però è come se alle origini del latino esistesse una duplicità di concezione del moto, che solo in epoca tarda si è convogliata in un comportamento unitario verso i satelliti, con un effetto in qualche modo dovuto alla creolizzazione, ma in maniera contraddittoria con quello che è l'esito romanzo.

Infatti i verbi che sono stati scelti nei vari esiti romanzi sembrano per lo più verbi di moto con direzione derivati dai verbi latini prefissati o sono verbi di nuova formazione sempre attraverso prefissazione, ormai non più apprezzabile e riconoscibile nelle lingue di arrivo (*ex-ire* > 'escire/uscire' *amb-ire* > \* *ambitare* > 'andare', *ab-ante* > 'avanzare', *passum* > 'passare', *ad ripam* > 'arrivare')<sup>21</sup>. Tant'è che è sempre necessaria la preposizione a reggere il nome per esprimere la direzione che è insita in quella che è ormai la radice verbale, ma non è più intesa come prefisso. Si pensi all'italiano in cui tutti verbi di moto con direzione ('andare', 'venire') necessitano della preposizione, che per esempio risulta ancora utile per distinguere: 'vengo da Roma', e 'vengo a Roma', 'vado a Roma' e 'vado da Roma a Latina'. Ci chiediamo fino a che punto l'occorrenza di necessarie preposizioni di direzione sia diversa dai satelliti. Rimarrebbe come aspetto distintivo dai verbi di moto con maniera il tratto semantico negativo per cui 'venire' o 'andare' non esprimerebbero la maniera del moto, o forse questa sarebbe neutralizzata perché in epoca antica essa veniva interpretata sempre come un sottinteso spostamento a piedi.

Per quanto riguarda l'italiano, inoltre, è stato recentemente mostrato da Jacobini (in stampa) che nella nostra lingua i verbi di moto con maniera sono numerosissimi (103 lemmi) e in quantità paragonabile a quelli dei lessici verbali delle lingue endocentriche germaniche (147 tedesco, 110 inglese) o slave (73 russo). La caratteristica sembra essere presente in maniera esclusiva nell'italiano, che in questo si differenzerebbe dalle altre lingue romanze (42 spagnolo, 61 francese). C'è poi un altro aspetto da segnalare, in italiano i verbi di moto con maniera possono comportarsi come quelli di direzione, ovvero se appropriatamente accompagnati da preposizioni, come richiedono necessariamente anche quelli di moto con direzione, possono esprimere la direzione ('nuotare da x a y', 'strisciare da x a y', 'balzerellare da x a y')<sup>22</sup>.

Che cosa significa? Forse una traccia di ricerca è da riportare alla situazione tutta particolare della derivazione diretta dell'italiano dal latino, ovvero alla

<sup>21</sup> Questa serie di esempi è ripresa da Herslund (2005).

<sup>22</sup> A questo proposito sembra da sfatare l'assunzione che il francese, che tra le lingue romanze è ritenuto il più radicale nella distinzione tra concezione del movimento con direzione o senza, non permetterebbe un uso misto di un verbo di moto con direzione più la maniera. Dovrebbero essere ammesse solamente costruzioni come *aller du bateau à la rive à la nage*, *aller de la classe à la maison en courant*, ma ormai sembrano entrati nell'uso parlato costrutti come *nager du bateau à la rive*, *courir de la classe à la maison*.

conservazione di una duplicità di concezione del movimento che sembra essere stata presente fino dalle origini nel latino e in qualche modo perpetuata nell'italiano. Del resto sappiamo che anche la derivazione dal latino è diretta ma duplice: da un lato da quel latino tardo che è all'origine della creolizzazione e dall'altro dal latino colto che porta alla conservazione e al recupero. I verbi di moto con maniera dell'italiano sono forse lemmi classici derivati dal latino, non tanto come derivazione diretta delle singole parole, ma come concezione del movimento, magari di solo uso letterario, per cui tale particolarità si potrebbe spiegare nella duplicità di rapporto con il latino del quale viene mantenuta anche la componente colta? Oppure sono lemmi totalmente nuovi e dunque appositamente concepiti in una strategia del movimento che in italiano, come caratteristica singolare nel campo romanzo ma concorde con il latino, è aperta sia al movimento con direzione che a quello con maniera? Evidentemente anche in questo caso ci sembrano necessarie ricerche quantitative su corpora che mostrino sia le frequenze sia le modalità (scritto / parlato) sia i registri (formale / informale / letterario) di uso in modo da evidenziare il valore dei diversi lessici verbali.

Esiste poi la valutazione dei correlati morfo-sintattici italiani che appare delicata se portata a confronto sia con il latino che con il danese, come esempi di lingue endocentriche. Infatti l'italiano, data la riduzione della flessione casuale e data la sua complessiva subordinazione analitica, da un lato può essere considerato più esplicito e più "semplice" del latino in un percorso di creolizzazione che porterebbe all'esocentrismo, d'altro canto però rispetto al danese mostra una maggiore strutturazione e complessità morfo-sintattica. Le correlazioni morfosintattiche che sono state evidenziate per le lingue endocentriche moderne, come il danese, con la preferenza per una strutturazione non ipotattica del periodo e lo scarso impiego della subordinazione con forme verbali infinite e delle nominalizzazioni, sembrano discostarsi totalmente dal latino, lingua ugualmente classificata come endocentrica sulla base del lessico verbale di moto. La sintassi latina evidentemente struttura in modo ipotattico e con subordinazione nominalizzata accordata per caso, che appare una delle forme di subordinazione più vincolante che una lingua possa prevedere, quindi sotto questo aspetto sarebbe al polo estremo rispetto al danese. Del resto ci sembra che un costrutto nominalizzato e retto da accordo di caso, come quelli del latino, sia abbastanza diverso da un costrutto di subordinazione con gerundio o participio, aggiunto in maniera sostanzialmente paratattica, come quelli dell'italiano. Quindi quando si parla di subordinazione fatta per nominalizzazione bisognerebbe distinguere. Più in generale, è possibile dare una valutazione assoluta della complessità sintattica di una lingua o questa è sempre relativa?

Le lingue possono essere classificate in maniera netta come esocentriche o endocentriche o dobbiamo ipotizzare una gradualità e variazione al loro interno?

Infine qual è l'identità dell'italiano? È un caso speciale dentro l'universo romanzo proprio per la sua derivazione insieme diretta e duplice dal latino?

## Bibliografia

- Alberti, L. B. 1996. «*Grammatichetta*» e altri scritti sul volgare, A cura di G. Patota, Roma: Salerno Editrice.
- Baron, I. 2005. Endocentric versus Exocentric languages: an obstacle to European Integration?. In I. Korzen & P. D'Achille (eds), *Tipologia linguistica e società*. Pisa: Franco Cesati Editore, 93-110.
- Bloomfield, L. 1933. *Language*. London: Allen & Unwin.
- British National Corpus (BNC). <http://www.natcorp.ox.ac.uk/>
- Cresti, E. 2000. *Corpus di italiano parlato*. Firenze: Accademia della Crusca.
- Cresti, E. 2009. Clitics and anaphoric relations in informational patterning: a corpus driven research in spontaneous spoken Italian (C-ORAL-ROM). In L. Mereu (ed.), *Information structures and its interfaces*. Berlin / New York: Mouton de Gruyter, 171-203.
- Dardi, A. 1992. *Dalla provincia all'Europa. L'influsso del francese sull'italiano tra il 1650 e il 1715*. Firenze: Le Lettere.
- Herslund, M. 2005. Lingue endocentriche e lingue esocentriche: aspetti storici del lessico. In I. Korzen & P. D'Achille (eds), *Tipologia linguistica e società*. Pisa: Franco Cesati Editore, 19-30.
- Iacobini, C. (in stampa). The number and use of manner verbs as a cue for typological change in the strategies of motion events coding. In G. Marotta & A. Lenci & L. Meini & F. Rovai (eds) *Proceedings del Convegno Internazionale "Space in Language"*, Pisa 8-10 ottobre 2009. Pisa: Pisa University Press.
- Izre'el, S. & Mendelson, M. 2002. The Emergence of Spoken Israeli Hebrew. In S. Izre'el (ed.), *Speaking Hebrew: Studies in the Spoken Language and in Linguistic Variation in Israel (Te'uda, 18)*. Tel Aviv: Tel-Aviv University, 217-238.
- Korzen, I. 2005a. Lingue endocentriche e lingue esocentriche: lessico, testo e pensiero. In I. Korzen & P. D'Achille (eds), *Tipologia linguistica e società*. Pisa: Franco Cesati Editore, 31-54.
- Korzen, I. 2005b. Struttura linguistica e schema cognitivo: tipologie a confronto. *Copenhagen Studies in Language* 31, I: 123-134.
- Meini, L. & McGillivray, B. (in stampa). Between semantics and Syntax: spatial verbs and preposition in Latin. In G. Marotta & A. Lenci & L. Meini & F. Rovai (eds) *Proceedings del Convegno Internazionale "Space in Language"*, Pisa 8-10 ottobre 2009. Pisa: Pisa University Press.
- Migliorini, B. 1961. *Storia della lingua italiana*. Firenze: Sansoni.
- Moneglia, M. 2010. Predicati generali ad alta frequenza nei corpora orali delle lingue romanze. In M. Iliescu, H.M. Siller-Runggaldier P. & Danler (eds), *Actes du XXV Congrès International de Linguistique et de Philologie Romanes* Innsbruck, 3-8 settembre 2007, Tome IV. Berlin / New York: Mouton de Gruyter, 497-506.
- Moneglia, M. & Panunzi, A. (in questo volume). I verbi generali nei corpora di parlato. Un progetto di annotazione semantica cross-linguistica.
- Panunzi, A. & Moneglia, M. 2004. La variazione primaria del verbo nel lessico dei corpora di parlato, in F. Albano Leoni, F. Cotugno, M. Pettorino & R. Savy (eds), *Atti del convegno nazionale "Il parlato Italiano"*. Napoli: M. D'Auria Editore, 1-24 (C4).

- Panunzi, A. 2010. *La variazione semantica del verbo essere nell'italiano parlato*. Firenze: Firenze University Press.
- Lessico Etimologico Italiano (LEI) dal 1979. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Rohlf, G. 1966-69. *Grammatica storica della lingua italiana e dei suoi dialetti*. Torino: Einaudi.
- Salvi, G. & Renzi, L. (ed.) 2010. *Grammatica dell'italiano antico*. Bologna: Il Mulino.
- Scarano, A. 2003. *Fraasi relative e pseudo-relative. Sintassi, semantica e articolazione dell'informazione*. Roma: Bulzoni.
- Slobin, D.I. 2004. The many ways to search for a frog. Linguistics typology and the expression of motion events. In S. Stromqvist & L. Verhoeven (eds) *Relating events in narrative: typological and contextual perspectives*. Mahwah (NJ): Lawrence Erlbaum Associates, 219-257.
- Tagliavini, F. 1949. *Le origini delle lingue neolatine*. Bologna: Patron.
- Talmy, L. 1985. Lexicalization patterns: semantic structure in lexical forms. In Th. Shopen (ed.), *Language Typology and syntactic description*. Cambridge: Cambridge University press, 57-149.
- Talmy, L. 2000. *Toward a cognitive semantics*. Cambridge MA: The MIT Press.
- Tomasin, L. 2010. *Storia linguistica di Venezia*. Roma: Carocci.



L'ITALIANO IN UNA PROSPETTIVA DI TREEBANK.  
IL COPENHAGEN DEPENDENCY TREEBANK PROJECT:  
ASPETTI SINTATTICI E TESTUALI

Iørn Korzen

Copenhagen Business School

**1. Il progetto delle “Copenhagen Dependency Treebank”, “CDT”<sup>1</sup>**

Parlando di “estensioni”, o sviluppi, di una tipologia linguistica, se ne possono intendere propositi più concretamente estensionali, intenzionati ad includere nuovi campi di studi, campi linguistici e non, oppure propositi che si potrebbero definire “intensionali”, riguardanti invece nuove metodologie, principi o approcci di studio. Stessa duplice interpretazione riguarda gli sviluppi dello studio comparativo delle lingue endo- ed esocentriche.

Questo volume contiene esempi sia di sviluppi estensionali della tipologia delle lingue endo- ed esocentriche (linguisticamente estensionali per esempio i contributi di Herslund e di H. Korzen, extralinguisticamente estensionali i contributi di Lundquist e di Smith), sia di sviluppi intensionali di tale tipologia: le seguenti pagine vogliono essere un esempio di quest'ultimo genere, il contributo di Müller ne è un altro.

Entrambi quest'ultimi contributi presentano l'applicazione di una nuova metodologia a due campi di studi delle lingue endo-/esocentriche, il campo morfologico (il contributo di Müller) e il campo discorsivo o testuale (le pagine seguenti). I due contributi sono fra i primi risultati della partecipazione di Müller e mia al progetto delle “Copenhagen Dependency Treebank”, “CDT”, con base alla Copenhagen Business School. Tale progetto ha lo scopo di creare delle treebank parallele di cinque lingue diverse, l'italiano, lo spagnolo, il tedesco, l'inglese e il danese, ed è in sé l'estensione di una treebank parallela danese-inglese creata dal

---

<sup>1</sup> Ringrazio Remo Stefano Chiari e Nora Galli de' Paratesi per utili suggerimenti e preziosi commenti ad una versione precedente di questo lavoro.

Cresti E., Korzen I. (eds), *Language, Cognition and Identity. Extensions of the endocentric / exocentric language typology* © 2010 Firenze University Press.

nostro collega, Matthias Buch-Kromann, e consistente di 474 testi con un totale di 100.200 parole.

I testi sono stati scelti a caso da un corpus danese chiamato “PAROLE”, che consiste di 1.553 testi di una vasta gamma di tipi, generi e argomenti diversi, provenienti da quotidiani, riviste, mass media, libri, ecc. per un totale di 250.000 parole, tutte annotate per categorie grammaticali<sup>2</sup>.

I testi scelti sono stati tradotti dal danese nelle altre lingue e annotati morfosintatticamente secondo principi descritti da Kromann, Lyng e Mikkelsen (2004), vedi anche la sezione 2 sotto; i testi danesi sono scaricabili dal sito dello stesso riferimento bibliografico, cfr. anche Kromann (2004a). Nel caso dei testi inglesi e danesi l’annotazione morfosintattica è già compiuta per tutte le 100.200 parole, per le altre tre lingue è una prima ambizione delle “CDT” arrivare a corpora annotati di 70.000 parole ognuna<sup>3</sup>. Oltre al collega Buch-Kromann, a Müller e al sottoscritto, cinque nostri laureati, tre ispanisti e due italianisti, partecipano o hanno partecipato al progetto, assunti come annotatori o assistenti di ricerca. Uno dei due italianisti è adesso legato al progetto come dottorando.

La creazione di treebank è tutt’altro che invenzione nuova, ma forse le “CDT” possono vantarsi di due particolarità: il numero di lingue coinvolte, cinque appunto, e il numero di livelli linguistici inclusi nell’annotazione. Oltre ai livelli grammaticale e sintattico (comprendente l’analisi logica delle frasi semplici e complesse), sono annotati gli altri due livelli linguistici cui ho accennato sopra: l’uno “superiore” alla frase, cioè il livello del discorso, ossia testuale, l’altro “inferiore” alla frase, cioè il livello morfologico nel caso di unità lessicali complesse, formate da radici più affissi e/o dalla composizione di più unità lessicali. Nelle pagine seguenti parlerò dei due livelli sintattico e testuale, con particolare riguardo a quest’ultimo ed al corpus italiano, mentre il livello morfologico è argomento di Müller (questo volume).

Fondamentale per il nostro progetto, che è in corso e relativamente giovane, è che esso ha come obiettivo un modello unificato dei sistemi di analisi e di annotazione dei tre livelli, e mentre soprattutto l’annotazione sintattica è un fenomeno ormai piuttosto conosciuto e diffuso (anche se le treebank italiane sono piuttosto poche)<sup>4</sup>, il che vale – sebbene in minor grado – anche per l’annotazione del

---

<sup>2</sup> Per più informazioni sul corpus PAROLE rimando a <http://korpus.dsl.dk/e-resurser/parole-korpus.php>, da cui il corpus è scaricabile.

<sup>3</sup> I testi tradotti in italiano, spagnolo, tedesco e inglese sono scaricabili da questo sito: <http://code.google.com/p/copenhagen-dependency-treebank/wiki/CDT>

<sup>4</sup> Per l’ISST, l’Italian Syntactic-Semantic Treebank: cfr. per esempio Montemagni et al. 2003a/b. Per la Turin University Treebank, cfr. <http://www.di.unito.it/~tutreeb/index.html>, e per la Venice Italian Treebank, cfr. <http://torvald.aksis.uib.no/corpora/2005-1/0385.html>.

discorso, presumiamo che l'unificazione dei tre livelli, nonché l'inclusione di cinque lingue diverse, siano proprietà uniche del progetto di Copenhagen.

## 2. L'annotazione sintattica delle CDT

Le treebank delle CDT sono state create con il programma “DTAG” (dove D sta per “dipendenza”) e con il DTAG annotation tool, entrambi ideati da Kromann (2004b) e scaricabili dal sito menzionato sotto tale riferimento bibliografico. L'annotazione consiste nel collegamento dei vari nodi attraverso relazioni di dipendenza espresse tramite l'uso di frecce arcuate. Le frecce sono etichettate con la relazione in questione e la loro direzione è *head* → *dependent*, cioè vanno dal nodo reggente al nodo dipendente, come illustrato nella figura 1 sotto. I nodi possono essere parole, parti di parole, gruppi di parole o segni di punteggiatura, e possono avere una o più frecce sia uscenti sia entranti. Le frecce sono disegnate o sopra i nodi (nel caso delle dipendenze “principali”) o sotto i nodi (nel caso delle dipendenze “secondarie”, cfr. di seguito). L'etichetta della funzione del nodo dipendente rispetto a quello reggente è indicata alla punta della freccia (cioè proprio sopra o sotto il nodo dipendente) nel caso di una sola freccia entrante e a metà freccia nel caso di più frecce entranti. I grafici sono generati in PostScript file leggibili con il PostScript viewer (Kromann 2004a; 2005), e il sistema è basato su command-line: per specificare per esempio una freccia dal nodo 4 al nodo 7, quest'ultimo con la funzione di *oggetto diretto* del nodo 4, l'utente digita il comando (inglese) “7 *doj* 4”.

L'annotazione sintattica è teoricamente basata sulla “Grammatica discontinua” di Buch-Kromann (2006: soprattutto cap. 2). Una dipendenza è definita *continua* se tutte le parole tra unità reggente e unità dipendenti si trovano all'interno dello stesso sintagma, e *discontinua* in altri casi (op. cit. 34). Nella dipendenza discontinua un'unità sintatticamente legata ad un sintagma o altro gruppo di parole si trova dunque al di fuori del sintagma o del gruppo di parole in questione (cfr. op. cit. 35-36).

Se c'è un verbo finito nell'unità sintattica annotata, esso funge da elemento reggente l'intera unità. La testa del sintagma nominale è il determinante e il nome (l'elemento lessicale) ne costituisce il *nominal object*, *nobj*; se non vi è determinante, il sostantivo fa da testa. La testa del sintagma preposizionale è la preposizione e un complemento nominale è *nobj*; la testa di una frase subordinata è la congiunzione, tranne che nella frase relativa, il cui verbo finito funge da testa. La figura 1 riporta l'annotazione dell'unità sintattica ““Sentivo che le cose non erano come avrebbero dovuto essere”, continuò lui”:

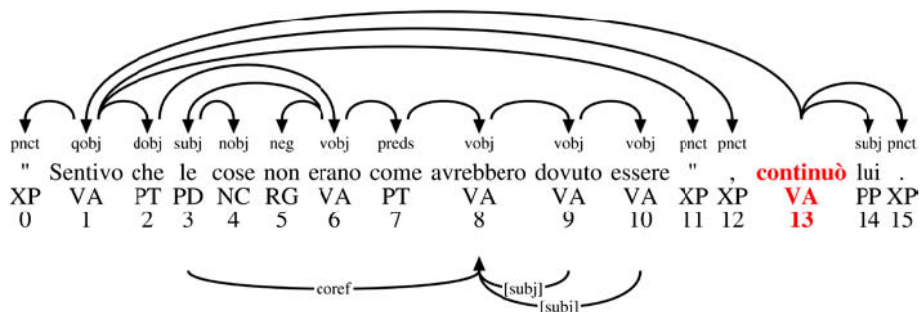


Figura 1. Annotazione sintattica di “*Sentivo che le cose non erano come avrebbero dovuto essere*”, *continuò lui*.

Una citazione funge da *quotational object*, *qobj*, del suo verbo dicendi, e il nodo più alto della citazione nella figura 1 è l’unità 1, etichettata quindi *qobj* dell’unità 13, nodo più alto dell’intera frase citata. La frase subordinata delle unità 2-10, il cui nodo più alto è la congiunzione, qui l’unità 2, funge da *oggetto diretto*, *dobj*, di ‘sentivo’, e il suo verbo finito, l’unità 6 ‘erano’, funge da *verbal object* della congiunzione. Il soggetto dell’unità 6 è il sintagma nominale ‘le cose’, retto dal determinante ‘le’, cui ‘cose’ funge da *nominal object*. Le parti infinite di un verbo complesso, come 9 e 10, fungono da *verbal objects* della parte verbale precedente.

Le virgolette e la virgola finale della citazione (*punctuation marks*, *pncnt*) sono tutte rette da essa, cioè dal suo nodo più alto, qui l’unità 1, e il punto finale della frase principale è retto dal nodo più alto di essa, cioè dall’unità 13.

Le relazioni menzionate finora ed annotate sopra i nodi sono considerate *principali*, mentre quelle annotate sotto i nodi sono quelle definite *secondarie*; quest’ultime includono per esempio le relazioni anaforiche (il soggetto zero dell’unità 8 è coreferenziale con le unità 3-4)<sup>5</sup> e il soggetto di forme verbali infinite (qui le unità 9 e 10 che hanno lo stesso soggetto dell’unità 8).

Tabella 1. Alcune dipendenze sintattiche delle “CDT”

COMPLEMENT RELATIONS	ADJUNCT RELATIONS
<b>subj</b> (soggetto)	<b>neg</b> (negazione)
<b>dobj</b> (direct object, oggetto diretto)	<b>attr</b> (attributivo)
<b>qobj</b> (quotational object, oggetto di citazione)	<b>cause</b> (avverbiale di causa)
<b>pobj</b> (prepositional object, oggetto preposizionale)	<b>time</b> (avverbiale di tempo)
<b>vobj</b> (verbal object, oggetto verbale)	<b>loc</b> (avverbiale di luogo)
<b>nobj</b> (nominal object, oggetto nominale)	<b>pncnt</b> (punctuation mark, segno di punteggiatura)

<sup>5</sup> Per l’annotazione delle relazioni anaforiche, cfr. Buch-Kromann, Korzen & Müller (2009), Buch-Kromann et al. (2010) e Korzen & Buch-Kromann (in stampa).

Come si è capito, le dipendenze sono indicate con i termini inglesi, e fra le più frequenti a livello sintattico si trovano quelle indicate nella tabella 1.

Come vedremo, alcune delle etichette della colonna a destra ricorrono nell'annotazione del livello testuale. Infatti, una delle differenze più importanti tra le lingue endocentriche (soprattutto scandinave) e le lingue esocentriche (romanze) è la maggiore complessità sintattica e testuale di quest'ultime. Come ho già dimostrato per esempio in Korzen (2003; 2007; 2009), le frasi di un testo italiano si distinguono, *ceteris paribus*, da quelle di un testo danese per la maggiore lunghezza e la maggiore complessità sintattica e informativa. Vi è una forte tendenza all'inclusione nello stesso periodo italiano di più proposizioni rispetto a quello che avviene in un periodo equivalente danese, dove invece si tende a dividere e ad accorciare le frasi. La tabella 2 dimostra la differenza di lunghezza dei periodi misurata come numero di parole<sup>6</sup> di una serie di tipi di testo diversi italiani e danesi<sup>7</sup>:

Tabella 2. Parole, periodi e parole per periodo in certi tipi di testo

Tipi di testi		Parole	Periodi	Parole per periodo
a. Testi giuridici	italiani	3000	95	31.6
	danesi	1690	84	20.1
b. Testi tecnici	italiani	4883	205	23.8
	danesi	4974	364	13.7
c. Newsgroups	italiani	4193	213	19.7
	danesi	1826	111	16.5
d. Siti web	italiani	4473	186	24.0
	danesi	3458	288	12.0
e. Testi narrativi	italiani	4050	187	21.7
	danesi	4592	220	20.9

Per questo motivo una proposizione per esempio causale o temporale, che in un testo italiano viene testualizzata come frase subordinata, eventualmente infinita, in una *adjunct relation*, può benissimo occorrere in un testo equivalente danese come frase

<sup>6</sup> Sono ben consapevole di tutte le riserve che bisogna fare con un calcolo basato sull'unità "parola"; comunque ritengo che le differenze, soprattutto dei testi giuridici, tecnici e dei siti web, siano convincenti.

<sup>7</sup> I testi giuridici della tabella 2 sono *la legge fallimentare* e *la legge sul divorzio*; quelli tecnici trattano la produzione di zucchero da barbabietola; i newsgroups (NUNC: <http://www.bmanuel.org/projects/ng-HOME.html>) trattano i due argomenti *caffè* e *dieta* e sono stati gentilmente messi a disposizione dal gruppo di ricerca coordinato da Carla Marellò e Manuel Barbera; i siti web provengono da quattro aziende produttrici di cioccolato, *Ferrero*, *Perugina*, *Toms* e *Sv. Michelsen*, e i testi narrativi provengono dal "Corpus di *Mr. Bean*", una serie di racconti di un filmato con il famoso comico inglese. Per più particolari e per riferimenti bibliografici più precisi rimando a Korzen (2009).

principale, sempre legata, comunque, al suo “nodo reggente” (per esempio un’altra frase principale) per la relazione causale o temporale che sia. Per questo motivo, l’inventario di relazioni sintattiche e testuali di un corpus che comprenda sia lingue endocentriche che lingue esocentriche, deve avere un certo numero di etichette identiche ai due livelli di analisi.

### 3. L’annotazione testuale delle CDT

#### 3.1 Fonti di ispirazione

Per la creazione di un modello di annotazione testuale ci siamo ispirati alle due grandi “scuole” dell’annotazione testuale: la cosiddetta Rhetorical Structure Theory (RST) che nacque verso la fine degli anni ’80 (cfr. per esempio Mann & Thompson 1987; Mann, Matthiessen & Thompson 1992; Matthiessen & Thompson 1988; Carlson, Marcu & Okurowski 2001; Taboada & Mann 2006a/b) e la Penn Discourse Treebank (PDTB) che è di data più recente (cfr. Prasad et al. 2007; 2008; Webber 2004; 2006). Questi due sistemi sono però monolingue, trattando solo l’inglese; unico altro esempio di paragoni testuali interlinguistici si trova in Mladová et al. (2008) (inglese-ceco), il quale è fortemente ispirato alla PDTB.

Una differenza fondamentale tra le treebank testuali menzionate e la nostra è che le prime si limitano all’annotazione testuale, ovvero a quella che la RST definisce come “clause combining”, dove per “clause” si intendono frasi sia principali che subordinate, e la PDTB si limita inoltre a frasi contigue. Invece il nostro sistema tratta tutti i casi di frasi subordinate al nostro livello sintattico, cfr. la figura 1, il che significa che al nostro livello testuale possiamo limitarci alle relazioni tra periodi ottenendo, forse, un quadro un poco più omogeneo.

Però in altri modi il nostro sistema ha approfittato sia della RST che della PDTB. Quest’ultima, a parte molte delle sue 32 relazioni<sup>8</sup>, ci ha particolarmente ispirato la distinzione gerarchica tra classe, tipo e sottotipo, cfr. l’esempio dato nella tabella 3 sotto, distinzione che permette una maggiore flessibilità agli annotatori: nei casi in cui un’analisi più dettagliata risulti incerta, possono semplicemente limitare l’analisi ad un livello superiore. La PDTB ci ha inoltre ispirato l’annotazione dei connettivi, sia quelli esplicitati, sia quelli impliciti ma inferibili. Tale annotazione si è dimostrata assai utile perché costringe l’annotatore, per così dire, ad argomentare ed a ragionare per l’analisi scelta. Vederne esempi nella sezione 3.3.1 sotto.

---

<sup>8</sup> Divise tra 29 relazioni semantiche e 3 non-semantiche, quest’ultime dette anche casi di “non-insertability of implicit connectives”, che sono i seguenti: la lessicalizzazione alternativa (“AltLex”), la coerenza basata su una entità (“EntRel”) e la mancanza di relazione interpretabile tra i periodi (“NoRel”). Cfr. Prasad et al. (2007, 22ss).

Anche la RST opera con una trentina di relazioni testuali diverse<sup>9</sup>, distinguendo fra l'altro tra relazioni *mononucleari*, o di *nucleo-satellite*, e relazioni *multinucleari* o di *lista* (Matthiessen & Thompson 1988: 288), relazioni di coordinazione retorica. Un *nucleo* contiene la parte più essenziale o saliente della sequenza testuale (in testi narrativi la parte che porta avanti la storia, in altri tipi di testo la parte che più chiaramente comunica l'intenzione della sequenza testuale), mentre i *satelliti* aggiungono informazione ausiliaria o secondaria e non funzionano comunicativamente – o funzionano male – senza il loro nucleo<sup>10</sup>. Dopo aver constatato la difficoltà della distinzione tra *nucleo* e *satellite* in termini di importanza o di salienza testuale, cfr. anche Stede (2008), abbiamo però deciso di abbandonare tale distinzione e di limitarci alla distinzione tra relazioni semantiche, che sussistono tra proposizioni, e relazioni pragmatiche, che sussistono tra atti linguistici (Stede op. cit. 325), con un piccolo gruppo di “altri fenomeni”, cfr. la sezione 3.2. tabella 3 cita come esempio la relazione semantica *time*:

Tabella 3. Un esempio della gerarchizzazione delle relazioni testuali delle CDT

Class	Main type	Subtypes	Possible connectives
Relazioni semantiche	TIME	TIME:cont[emporaneity]	<i>Mentre, Finché</i>
		TIME:pre[cedence]	<i>Prima, Tre giorni prima</i>
		TIME:post [succession]	<i>Dopo, Poco tempo dopo</i>

Come risulta dalla tabella 3, i tipi principali di relazione sono indicati con maiuscole e i sottotipi con minuscole dopo due punti.

I sistemi della PDTB e della RST sono stati fonti ispiratrici anche di altre treebank testuali; come accennato quella inglese-ceca si basa sulla PDTB, e il sistema proposto da Carlson, Marcu & Okurowski (2001) adotta il modello della RST. Carlson, Marcu & Okurowski operano però con ben 78 relazioni diverse suddivise in 16 classi; delle 78 relazioni, 53 sono mononucleari e 25 multinucleari. Molto più modesti sono invece Wolf & Gibson (2005) che, basandosi sulle relazioni testuali introdotte da Hobbs (1985), si accontentano di 11 *coherence relations*, e Stede (2008) che opera con 12 *conjunctive relations*, ma in cambio introduce altri tre livelli di analisi: coreferenza (relazioni anaforiche), struttura tematica e struttura

<sup>9</sup> Cfr. Taboada & Mann (2006a: 437): “As for the number of relations, the original list of 24 relations, expanded to 30 on the RST website [...] could have been shorter [...]. For various reasons, the list could not have been radically longer and still effective. Perhaps the strongest limit on distinguishing various relations in analysis is the possibility of distinguishing one relation from another, sometimes called *observability*.”

<sup>10</sup> “Mononuclear relations hold between two spans and reflect the situation in which one span, the *nucleus*, is more salient to the discourse structure, while the other span, the *satellite*, represents supporting information. Multinuclear relations hold among two or more spans of equal weight in the discourse structure.” (Carlson, Marcu & Okurowski 2001: 3).

intenzionale, quest'ultima divisa tra atti linguistici e relazioni pragmatiche, cfr. anche sotto.

Il sistema delle CDT si distingue però in altri modi sia dalla PDTB che dalla RST. Laddove la PDTB considera unicamente le relazioni tra unità contigue, la nostra unificazione con la struttura ad albero del livello sintattico ci permette di indicare casi in cui per esempio una relazione testuale non sussista tra segmenti contigui o tra i nodi più alti dei segmenti in questione. Un buon esempio di questo tipo si ha nel seguente:

- (1) Dopo la condanna, i due alti dirigenti della 6. juli Banken, il presidente del consiglio d'amministrazione Niels Johan Leer e il direttore John Olsdorff hanno immediatamente fatto ricorso contro la sentenza emessa dal tribunale di primo grado di Copenaghen chiedendo l'assoluzione. Anche il pubblico ministero si è riservato la possibilità di un ricorso.

Il presidente Leer è stato condannato ad un anno e mezzo di carcere e al pagamento di una multa di un milione di corone per abuso di mandato [...]. Il direttore Olsdorf è stato condannato a 6 mesi di carcere e al pagamento di una multa di 90.000 corone. (0531)

Nel caso citato il secondo paragrafo, 'il presidente Leer...' non si aggiunge al periodo precedente, invece elabora ed approfondisce il sintagma nominale 'la sentenza emessa dal tribunale di primo grado di Copenaghen' del primo periodo del brano citato. Nel nostro sistema ciò viene indicato con una freccia che va dal nodo più alto del sintagma nominale, ossia dal determinante 'la', al nodo più alto della frase 'Il presidente Leer...', ossia al verbo ausiliare finito 'è', con l'etichetta 'CONJ:elab', cioè tipo: CONJUNCT (congiunzione) e sottotipo: elaborazione. Invece la frase 'Anche il pubblico ministero...' si aggiunge alla frase precedente con la funzione di 'CONJ:add', cioè congiunzione:addizione; essa non elabora il segmento precedente, bensì vi si aggiunge con una nuova informazione. Torneremo su questo brano nelle sezioni 3.3.1 e 4 e nella figura 3.

Laddove la PDTB non annota relazioni testuali implicite tra un paragrafo e un altro né relazioni implicite intrafrasali, le CDT includono pure quelle. Le relazioni tra paragrafi sono etichettate con il segno '+' (vedi sezione 3.3.1, figura 2, unità 3, e sezione 4, figura 3, unità 32) e tutte le relazioni intrafrasali sono analizzate ed annotate al nostro livello sintattico.

Infine il fatto che le CDT includono lingue diverse dall'inglese, non solo lingue appartenenti allo stesso ceppo germanico come il tedesco e il danese, ma pure lingue appartenenti al ceppo romanzo, ci ha spinti, spesso costretti, ad operare con sottotipi particolari, a volte a causa di particolari inventari di connettivi, come vedremo nelle sezioni seguenti.



### 3.2 Le relazioni testuali delle CDT

Per l'etichettatura delle relazioni, le CDT si sono ispirate – oltre che alle fonti soprammenzionate – alla struttura “qualia” di Pustejovsky (1995: 67ss., 85ss.). Come è noto, per ogni unità lessicale Pustejovsky distingue quattro qualia o “ruoli”:

- il ruolo costitutivo (“constitutive quale”): la relazione tra l'entità e le sue parti costitutive (ad es. materiale, elementi/parti, contenuto),
- il ruolo formale (“formal quale”): che distingue l'entità all'interno di un dominio più ampio (ad es. forma, dimensione, posizione, colore),
- il ruolo agentivo (“agentive quale”): eventi legati alla creazione dell'entità,
- il ruolo telico (“telic quale”): scopo, funzione o risultato dell'entità.

Questi quattro qualia, o ruoli, sono anche pertinenti nelle relazioni semantiche: come vedremo in 3.2.1, un segmento testuale può svolgere appunto uno di questi ruoli, per esempio

- descrivere o precisare la “costituzione” (contenuto, elementi o parti) dell'entità cui si aggiunge (che può essere sia del primo, sia del secondo ordine nella terminologia di Lyons 1977: 442ss.), parliamo qui della relazione “costitutiva”,
- descrivere la forma o dimensione dell'entità, parliamo allora della relazione “formale” con i due sottotipi “descrittivo” (neutrale, oggettivo) e “valutativo” (parziale e/o soggettivo),
- esprimere la causa o la ragione dell'entità, parliamo della relazione “agentiva”, oppure
- indicare lo scopo o la conseguenza dell'entità, parliamo qui della relazione “telica”<sup>11</sup>.

Con in mente tutti i sistemi e le categorie menzionati abbiamo proceduto empiricamente, cioè indagando semplicemente i corpora a nostra disposizione. Per questo lavoro avevo formato un piccolo sottogruppo della nostra équipe, costituito da me stesso e da due laureandi, uno di italiano e l'altra di spagnolo, l'italianista ormai dottorando e l'ispanista assunta come assistente di ricerca. Dopo circa un anno di lavoro e circa 200 testi annotati, crediamo di esserci avvicinati ad un modello utile che sembri includere tutte le relazioni che abbiamo riscontrato. Va detto subito che siccome un discorso consistente di più periodi, come si sa, quasi inevitabilmente apre la possibilità di varie interpretazioni, anche l'annotazione testuale in molti casi rimane incerta e ambigua, e il nostro modello ne prende atto, come vedremo nella sezione 3.2.4 sotto.

---

<sup>11</sup> La struttura qualia è inoltre molto pertinente per le anafore associative, cfr. Korzen (2004).

Come si è detto sopra, il nostro modello opera con la distinzione di tre classi principali di relazioni:

- relazioni semantiche;
- relazioni pragmatiche;
- altri fenomeni.

In tutti i casi le frecce arcuate connettono i nodi più alti dei due segmenti in questione (che, ricordiamo, possono essere periodi, frasi, sintagmi o singole parole). Laddove vi è una chiara differenza di importanza o salienza testuale, come per esempio nelle relazioni concessive o nelle relazioni di “scena” (cfr. gli esempi (7) e (19) sotto), la freccia va dal segmento principale al segmento concessivo o di scena. In altri casi la freccia va semplicemente da sinistra verso destro, seguendo l’evoluzione del testo. Cfr. le sezioni seguenti.

### 3.2.1 *Relazioni semantiche*

Le relazioni definibili semanticamente sono 10 con un totale di 23 sottotipi. Seguedo il modello della tabella 3 cito nella pagina a fianco (tabella 4) l’inventario con esempi di connettivi italiani.

Seguono alcuni esempi delle relazioni citate, provenienti dal corpus italiano delle CDT, corpus, si ricorda, di testi tradotti dal danese. Il segmento testuale che si aggiunge con la relazione semantica indicata (più frequentemente il segmento a destra che si aggiunge a quello a sinistra) appare in corsivo; esso è anche punto d’arrivo della freccia, e la relazione semantica è indicata alla punta della freccia; vederne un esempio nella sezione 4, figura 3. Un connettivo implicito ma inferibile è aggiunto tra parentesi quadre, mentre un connettivo esplicitato è sottolineato.

- (2) CONST:rest  
 Non è previsto che Hafnia abbia troppa voce in capitolo. Si cerca di non far scappare i futuri partner della fusione, che potrebbero temere di essere sottomessi al controllo di altri. *Oppure, citando le parole di Preben Nygaard: “Se Hafnia detiene la maggioranza, potrebbe far scappare altri soggetti.* (0150)
- (3) FORMAL:descr  
 “Andiamo a casa tua” disse lei. *Era diversa da come me l’ero immaginata.* (0602)

In un caso come (3), la freccia va da ‘lei’ (il nodo che viene descritto, ma non il nodo più alto di quella frase) a ‘Era’.

Tabella 4. Le relazioni semantiche delle CDT

Main type	Subtypes	Possible connectives
CONST[itative]	CONST:apart [part-of relation]	<i>Incluso, Tra cui</i>
	CONST:elab[oration]	
	CONST:exem[plification]	<i>Per esempio</i>
	CONST:rest[atement]	<i>In altre parole, Cioè</i>
FORMAL	FORMAL:descr[ription: objective and/or neutral description]	
	FORMAL:eval[uation: personal and/or subjective, positive/negative]	
AGENTIVE	AGENTIVE:reas[on]	<i>Perché, Dato che</i>
	AGENTIVE:expl[anation]	<i>Infatti</i>
	AGENTIVE:sbj [a personal argument to support a claim]	<i>Perché, Dato che, Infatti</i>
TELIC	TELIC:goal	<i>Per</i>
	TELIC:cons.dir[ect] (physical, objectively observed consequence)	<i>Quindi, Perciò</i>
	TELIC:cons.sbj (subjective conclusion on behalf of the speaker)	<i>Quindi, Perciò</i>
COND[ition]	–	<i>Se, A condizione che</i>
CONC[ession]	–	<i>Anche se, Sebbene</i>
TIME	TIME:cont[emporaneity]	<i>Contemporaneamente</i>
	TIME:pre[cedence]	<i>Prima, Tre giorni prima</i>
	TIME:post [temporal succession]	<i>Dopo, Poco tempo dopo</i>
CONJ[unction]	CONJ:add[izione]	<i>E, Oltre a ciò</i>
	CONJ:elab[orazione]	
	CONJ:seq[uenza]	liste di eventi, risultati sportivi, ricette ecc.
CONTR[ast]	CONTR:dir[ect] [contrasto tra segmenti testuali esplicitati]	<i>Ma, Invece</i>
	CONTR:sbj [subjective] [contrasto tra segmento esplicitato e segmento soggettivamente inferito]	<i>Ma, Invece</i>
DISJ[unction]	DISJ:dir[ect] [alternativa tra segmenti testuali esplicitati]	<i>Oppure, Altrimenti</i>
	DISJ:sbj [subjective] [alternativa tra segmento esplicitato e segmento soggettivamente inferito]	<i>Oppure, Altrimenti</i>

- (4) AGENTIVE:reas  
 Simonsen non si sbilancia troppo nel commentare le future strategie d'investimento. [perché] *Egli è ben consapevole che qualsiasi mossa verrà seguita con attenzione dal mercato azionario.* (0781)
- (5) TELIC:cons.dir  
 Tra tre anni un grande numero di traghetti del sistema nazionale danese dovrà cercare dei nuovi mari: [perciò] *una vicenda emozionante entrerà nella storia.* (0848)

In un caso come (5), l'annotazione delle CDT indica pure che il confine tra le frasi è segnato dai due punti.

- (6) COND  
 “Ma dopo l'operazione al ginocchio non sentirà più dolore. A condizione però di effettuare allenamenti specifici dei muscoli delle cosce” sostiene il chirurgo ortopedico. (1270)
- (7) CONC  
 [ammetto che] *Lo so che non ci conosciamo da molto, ma non è importante. Vorresti...vorresti sposarmi, Trine?* (0620)
- (8) TIME:post  
 Il Ministero degli Interni prevede che questa parte del lavoro si concluda prima del 1. giugno 1984. Subito dopo, il Ministero e la Difesa Civile valuteranno l'eventualità di far proseguire i lavori. (0596)
- (9) CONJ:add  
 Il direttore della banca è stato condannato a 6 mesi di carcere. [e] *Due membri del consiglio d'amministrazione sono stati assolti.* (0531)<sup>12</sup>
- (10) CONJ:elab  
 Era diversa da come me l'ero immaginata. *Nel taxi appoggiò il suo capo sulla mia spalla.* (0602), il seguito dell'esempio (3)

Va aggiunto che in molti casi può essere difficile distinguere tra CONJ:add e CONJ:elab, ragion per cui l'annotatore può decidere di rimanere al livello di tipo principale, CONJ.

---

<sup>12</sup> Le relazioni CONJ, CONTR e DISJ sono definite multinucleari, o di lista, nella tradizione della RST.

## (11) CONTR:dir

Le nuove aziende dovranno essere collocate soprattutto a sud del centro fiere di Bella Center, e cioè creare una zona che si orienta a diventare decisamente urbanizzata. *Si darà invece maggiore spazio alla natura a nord del Centro, destinando 31 ettari come area verde.* (1420)

## (12) CONTR:subj

È del tutto evidente che ci sarà un aumento degli investimenti sul versante estero nei prossimi anni. *Ma noi non operiamo imprudentemente, per cui anche nei prossimi anni si compreranno azioni di società straniere con filiali danesi.* (0781)

Il contrasto “diretto” sussiste tra due segmenti testuali espliciti, mentre il contrasto “soggettivo” sussiste tra un segmento testuale esplicito ed uno implicito ma inferibile<sup>13</sup>. Nel caso di (12) il segmento inferito in contrasto con il secondo periodo è qualcosa come: ‘Tali investimenti potrebbero definirsi operazioni imprudenti’.

## (13) DISJ:dir

Siamo sicuri che quei batteri o virus non inizino a creare disordini negli ecosistemi naturali? *Oppure ad arrecare danni all'uomo e agli animali?* (0215)

## 3.2.2 Relazioni pragmatiche

Certi segmenti testuali si distinguono per una funzione illocutoria diversa da quella del co-testo, per cui abbiamo a che fare con relazioni *pragmatiche* nel senso di Stede (2008). Può trattarsi di comandi, di esclamazioni, di domande, di risposte, ecc., e oltre a tale funzione vi può essere o non essere, con il co-testo precedente, una relazione semantica dei tipi trattati in 3.2.1, per esempio una relazione di elaborazione o di conseguenza, nel qual caso tale relazione viene aggiunta all’etichetta illocutoria con il segno ‘&’, cfr. la sez. 3.2.4 sotto.

Simile alla funzione di tali segmenti è quella dei segnali interazionali, per esempio di attenzione o di interruzione<sup>14</sup>, nonché i segmenti che eseguono la funzione di consolidamento o rafforzamento. Può trattarsi del consolidamento di un segmento del co-testo (ossia di un’azione verbale), cioè di una specie di *fonte di autorevolezza* a favore dell’uso di un’espressione o elemento testuale, oppure del consolidamento di un’azione fisica eseguibile dall’interlocutore, cioè di una *motivazione* o di uno *strumento* che lo rende capace di eseguire l’azione in questione. In tutti i casi il segmento in questione si distingue dal co-testo per l’intenzione e/o per l’atto illocutorio del parlante:

<sup>13</sup> La PDTB usa qui il termine *contrasto pragmatico*: Prasad et al. (2007: 33).

<sup>14</sup> Da Bazzanella (1994: 156ss) e (2001: 179ss) chiamati *segnali discorsivi*, da Skytte (1999°: 299) e (1999b: 464ss) chiamati *segnali interazionali*.

Tabella 5. Le relazioni pragmatiche delle CDT

Main type	Subtypes	Segment types
DIREC[tive]		costrutti all'imperativo o all'infinito
EXPR[essive]		<i>Mi dispiace! Le mie condoglianze!</i>
QUEST[ion]		
ANSW[er]		
INTACT	INTACT:attn [attention]	<i>Ah; Veramente; Sì?</i>
[interaction signals]	INTACT:inter[rupcion]	<i>Ma..., Ma</i>
	CONSOL:source	
CONSOL[idation]	CONSOL:motiv[ation]	<i>I prezzi non sono mai stati così bassi.</i>
	CONSOL:inst[rument]	<i>Per il catalogo scrivete a...</i>

Negli esempi seguenti i segmenti in corsivo indicano, come sopra, il punto d'arrivo delle frecce e quindi i segmenti con la funzione indicata. La domanda può essere seguita o no da una risposta:

- (14) DIREC  
 “Penso che siamo stati bene, no? *Mi raccomando, non dire cose false su di noi!*”  
 (0438)
- (15) QUEST  
 Ora dovrà vivere diversi mesi con il terrore di essere stata contagiata dall'AIDS. *Cosa pensi che significhi per una bambina?* (0096)
- (16) ANSW  
 Ci si può lasciar sedurre nel parco di Kongens Have con 35 gradi, sdraiati sotto l'ombra di un albero, in più completamente vestiti, bevendo acqua, fumando sigarette e ascoltando le confessioni di un provocatore pop?  
*Sì!* (1428)
- (17) INTACT:attn  
 – “C'è una cosa che ti vorrei chiedere.”  
 – “Sì?” (0620 – brano che precede quello citato in (7) sopra e in (20) sotto)
- (18) CONSOL:source  
*Uno fra i maggiori esperti legali del paese per le forme di coabitazione, è l'avvocato Svend Trangeled. Egli afferma che soprattutto nelle abitazioni in cooperativa si possono annidare le insidie maggiori.* (1035)

### 3.2.3 Altri fenomeni

Infine restano tre fenomeni, diversi tra loro e diversi da quelli menzionati finora. I primi due riguardano la strutturazione formale del testo: l'etichetta *scene* è usata per titoli, intestazioni e sim. che presentano la scena del testo che segue, e l'etichetta

*repaired* indica brani e segmenti (soprattutto di testi orali o orali scritti) incompiuti e completati nel co-testo seguente. Il terzo fenomeno, *joint* (termine proposto dalla RST), riguarda invece i casi in cui due segmenti testuali si susseguono senza una chiara relazione semantica, pragmatica o diversamente funzionale l'uno con l'altro:

Tabella 6. Altri fenomeni

Relation types	Segment types
SCENE	titoli, intestazioni
REP[aired]	segmenti testuali incompiuti
JOINT (nessuna relazione semantica, retorica o funzionale evidente)	

## (19) SCENE

*Villa, aereo e automobili*

Il gruppo degli azionisti è amareggiato dal fatto che Klaus Riskær in qualità di presidente del Consiglio di amministrazione possa vivere in una villa da 27 milioni, avere un aereo e diverse automobili a disposizione [...]. (0178)

In tali casi la freccia va dal nodo più alto del testo che segue al titolo (qui: 'è [amareggiato]') al titolo stesso (qui 'Villa, aereo e automobili').

## (20) REP

Lo so che non ci conosciamo da molto, ma non è importante. *Vorresti... vorresti sposarmi, Trine?* (0620)

Nei casi di "REP" la freccia va dal segmento *riparante* (qui: 'Vorresti sposarmi, Trine?') al segmento *riparato* (qui: 'Vorresti...').

## (21) JOINT

Abbiamo appetito, ma in tutti i ristoranti ci respingono. *In una strada laterale notiamo un neon verde che lampeggia.* (0235)

Nell'esempio citato la connessione tra le due frasi viene chiarita dal co-testo seguente: 'Qui potremo avere da mangiare, ma dobbiamo pagare per entrarci.' Tuttavia la relazione immediata tra le due frasi citate non risulta chiara e viene quindi etichettata *joint*.

3.2.4 *Casi incerti o ambigui*

È ben noto che la lingua è sottospecificata rispetto alla realtà che vuole descrivere, ed è altrettanto ben noto che spesso la relazione tra due periodi di un testo risulta o incerta o ambigua, cioè aperta a più di una interpretazione. L'annotazione delle CDT prende atto di questo e permette l'aggiunta del punto interrogativo "?" nel caso di

incertezza di una relazione e l'uso della lineetta verticale “|” nel caso di incertezza tra due relazioni, come nel seguente esempio:

(22) CONJ/(e)|TIME:post

I dieci della Superliga devono incontrarsi due volte nel corso della primavera, per un totale di 18 giorni che termineranno il 23 giugno, quindi il programma sarà il seguente: [...]

*Cosa succederà dopo le vacanze estive, l'Associazione della Divisione calcio lo deciderà con l'assemblea generale del 26 gennaio.* (0043)

In (22) il periodo in corsivo può essere interpretato o come congiunto al periodo precedente (con il connettivo implicito ‘e’) o come indicatore di eventi successivi a quelli descritti nel primo periodo.

In altri casi la relazione tra due periodi si può prestare a più di un'interpretazione, per esempio perché il secondo segmento contiene un connettivo indicatore di una relazione particolare mentre l'interpretazione di un'altra rimane altrettanto possibile, oppure perché il secondo segmento svolge una funzione illocutoria particolare (cfr. la sezione 3.2.2) e allo stesso tempo vi è una relazione semantica, per esempio di elaborazione o di conseguenza, tra i due segmenti. In tali casi le CDT si servono del segno “&”; in (23) il contenuto del periodo in corsivo può essere considerato sia come una conseguenza diretta di quello precedente dato il connettivo ‘quindi’, sia come una semplice descrizione della galleria in questione, e in quest'ultimo caso sia la descrizione valutativa ‘FORMAL:eval’, sia quella oggettiva ‘FORMAL:descr’, possono applicarsi per gli aggettivi ‘eleganti’ e ‘pochi’ rispettivamente:

(23) TELIC:cons.dir/quindi&FORMAL:descr|eval

La galleria d'arte continua ad avere poco spazio espositivo, nonostante il recente trasloco in locali più ampi su Gothersgade; neanche qui lo spazio consente di esporre molto, con la conseguente ovvia limitazione dei mobili in mostra. *Quindi gli articoli sono pochi, ma eleganti.* (0159)

### 3.3 Ulteriori annotazioni

Oltre alle relazioni stesse, che arrivano ad un totale di 19 tipi e 35 sottotipi (quindi più o meno della stessa dimensione della RST e della PDTB), le annotazioni delle CDT includono altri fenomeni ancora.

#### 3.3.1 *Il connettivo*

Come accennato in 3.1 sopra, il nostro modello comprende anche, se possibile, l'annotazione del connettivo che precisa la relazione tra i periodi. Se un tale



connettivo è esplicitato nel testo, come negli esempi (2), (6), (8), (11), (12), (13) e (23) sopra, esso viene aggiunto dopo la barra nella lingua del testo annotato:

- (2): CONST:rest/oppure
- (6): COND/a condizione di
- (8): TIME:post/subito dopo
- (11): CONTR:dir/invece
- (12): CONTR:subj/ma
- (13): DISJ:dir/oppure
- (23): TELIC:cons.dir/quindi

Se non c'è un connettivo, ma uno è interpretabile, esso viene aggiunto tra parentesi, come in (4), (5), (7), (9):

- (4): AGENTIVE:reas/(perché)
- (5): TELIC:cons.dir/(perciò)
- (7): CONC/(ammetto che)
- (9): CONJ:add/(e)

Per illustrare con alcuni esempi torniamo brevemente al brano citato in (1), che si ripete qui con in grassetto i nodi più alti dei segmenti collegati:

- (1) Dopo la condanna, i due alti dirigenti della 6. juli Banken, il presidente del consiglio d'amministrazione Niels Johan Leer e il direttore John Olsdorff **hanno** immediatamente **fatto ricorso**<sub>0</sub> contro **la sentenza**<sub>1</sub> emessa dal tribunale di primo grado di Copenaghen chiedendo l'assoluzione. Anche il pubblico ministero **si è riservato**<sub>2</sub> la possibilità di un ricorso.

Il presidente Leer **è stato condannato**<sub>3</sub> ad un anno e mezzo di carcere e al pagamento di una multa di un milione di corone per abuso di mandato [...]. Il direttore Olsdorf **è stato condannato**<sub>4</sub> a 6 mesi di carcere e al pagamento di una multa di 90.000 corone. (0531)

L'annotazione testuale di questo brano avrebbe la struttura illustrata nella figura 2:

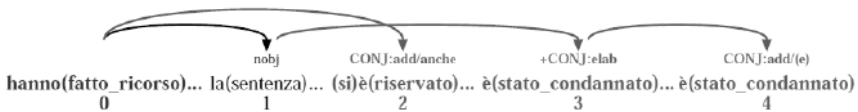


Figura 2. Annotazione testuale del brano citato in (1)

Per la funzione del sintagma nominale 1 come *nobj*, cfr. la sezione 2. Alcune relazioni sono caratteristiche per non avere particolari connettivi; ciò vale per FORMAL, CONST:elab, nonché per i tipi elencati nelle Tabelle 5 e 6.

### 3.3.2 *L'attribuzione*

Il fenomeno di *attribuzione* di discorso diretto o indiretto, pensieri, idee ecc. ad una o più persone, viene analizzato in modi diversi nei vari sistemi. Laddove la RST non include l'attribuzione tra le relazioni applicate, Carlson & Marcu (2001: 7ss) e Carlson, Marcu & Okurowski (2001: 5), che altrimenti adottano il modello della RST, la trattano come una relazione vera e propria. La stessa cosa si osserva in Wolf & Gibson (2005: 254)<sup>15</sup>, mentre La PDTB arriva ad elaborare un sistema particolare e piuttosto complesso di distinzioni e sottodistinzioni per la sua annotazione, sistema che include

- la fonte (che può essere la stessa del co-testo, diversa dal co-testo o generica),
- il contenuto (asserzione, opinione, dato di fatto o eventualità),
- una eventuale negazione (aggiunta al verbo dicendi in un processo di *raising*) e
- determinatezza (il fatto che l'implicazione dell'attribuzione può risultare indeterminata a causa di un co-testo negato o condizionato) (cfr. Prasad et al. 2006).

Nelle CDT abbiamo deciso di includere il fenomeno dell'attribuzione applicandovi però un sistema molto più elementare, che consiste semplicemente nell'aggiunta di una barra seguita dall'etichetta *ATTR* dopo il resto dell'annotazione. In caso di attribuzione di discorso o sim. a più persone viene aggiunto un numero per ogni persona: */ATTR1*, */ATTR2*, */ATTR3* e così via. Quindi in un caso del tipo

- (24) “A che ora vieni?” domandò Luigi.  
 “Alle 8” rispose Anna,

avremmo una freccia dal nodo superiore del primo periodo, ‘domandò’, a quello superiore del secondo periodo, ‘rispose’, con il numero legato alla persona ‘Anna’: il numero indica sempre la persona della battuta d'arrivo della freccia. L'aggiunta di ‘ATTR’ viene fatta sia per le frecce che collegano verbi dicendi, come in (24), sia per i casi in cui essi manchino, per cui le frecce collegano direttamente le parti del discorso riportato.

---

<sup>15</sup> Diversamente da Hobbs (1985) su cui gli studiosi si basano e con una motivazione che potrebbe sembrare un poco arbitraria: “Hobbs (1985) does not include an *attribution* relation. However, we decided to include *attribution* as a relation because [...] the texts we annotated are taken from news corpora. There, attributions can be important carriers of coherence structures.” (loc. cit.).

In questi casi, molto spesso la relazione tra i verbi dicendi non è quella più importante, lo è invece quella tra i brani di discorso citato. Il nostro sistema prende atto di questo, e l'etichetta di una relazione tra verbi dicendi, qui 'domandò' 'rispose', indica la relazione tra le battute, qui la relazione ANSW, cfr. l'esempio (16), non la relazione tra i due verbi dicendi (che è quasi sempre una semplice congiunzione). Un asterisco indica che tale relazione va trovata ad un livello sintattico "più basso", ossia tra i *quotational objects* (cfr. la sezione 2, figura 1) dei due verbi dicendi. L'asterisco viene aggiunto davanti all'etichetta se il livello più basso riguarda il segmento testuale a sinistra e dopo l'etichetta se invece riguarda il segmento a destra. In un caso come (24), dove la relazione in questione, ANSW, va trovata ad un livello più basso di entrambi i verbi dicendi, vanno aggiunti due asterischi, per cui l'etichetta della relazione 'domandòrispose' sarebbe \*ANSW/ATTR2\*. Contemporaneamente viene annotata – automaticamente – una relazione secondaria con una freccia sotto le parole (come nel caso delle relazioni *coref* e [*subj*] della figura 1), freccia che va da 'vieni' ad 'Alle', etichettata ANSW/ATTR2, cioè con la stessa etichetta della relazione principale ma senza gli asterischi.

#### 4. Osservazioni conclusive

Le Copenhagen Dependency Treebanks sono ancora alle prime armi e non è detto che il lavoro futuro non ci porti a rivedere l'inventario delle nostre relazioni di dipendenza testuali e, forse, sintattiche. Come accennato alla fine della sezione 1, forse il contributo più importante delle CDT è quello di unificare i tre livelli di annotazione, il livello morfologico, sintattico e testuale, e forse in tal modo di arrivare a definire e a descrivere un segmento testuale come una lunga relazione di dipendenza dal livello più alto testuale a quello più basso morfologico, nonché delle correlazioni e dei parallelismi tra i diversi strati linguistici. Inoltre tale procedura ci permette una maggiore precisione per esempio in casi in cui una relazione testuale non sussista tra i nodi superiori dei due segmenti attigui, ma tra nodi diversi, come si è visto nell'esempio (1) delle sezioni 3.1 e 3.3.1. La figura 3 qui sotto riporta la piena annotazione sintattica e testuale del brano citato in (1), per questioni di spazio con piccole omissioni del testo; le frecce più sottili indicano le dipendenze sintattiche e quelle più spesse le dipendenze testuali:

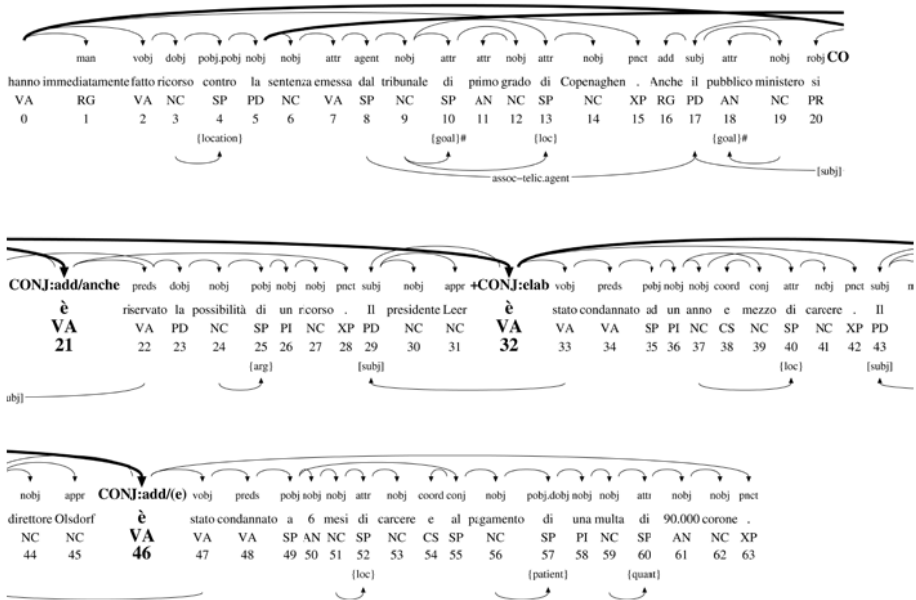


Figura 3. Annotazione sintattica e testuale del brano citato in (1)

Per le varie relazioni secondarie (semantiche, sintattiche ed anaforiche), annotate sotto il testo, rimando al manuale delle Copenhagen Dependency Treebanks (Buch-Kromann et al. 2010), scaricabile dall'indirizzo: <http://copenhagen-dependency-treebank.googlecode.com/svn/trunk/manual/cdt-manual.pdf>, nonché a Korzen & Buch-Kromann (in stampa).

### Bibliografia

Bazzanella, C. 1994. *Le facce del parlare. Un approccio pragmatico all'italiano parlato*. Milano: La Nuova Italia.

Bazzanella, C. 2001. I segnali discorsivi. In L. Renzi, G. Salvi & A. Cardinaletti (eds), *Grande grammatica italiana di consultazione*. Bologna: Il Mulino, 225-257.

Buch-Kromann, M. 2006. *Discontinuous Grammar. A dependency-based model of human parsing and language learning*. Tesi di dottorato, Copenhagen Business School.

Buch-Kromann, M. & Korzen, I. 2010. The unified annotation of syntax and discourse in the Copenhagen Dependency Treebanks. In *Proceedings of the Fourth Linguistic Annotation Workshop, ACL 2010*, Uppsala (Sweden) 15-16 July 2010. Taberg, 127-131.

Buch-Kromann, M., Korzen, I. & Müller, H.H. 2009. Uncovering the 'lost' structure of translations with parallel treebanks. *Copenhagen Studies in Language* 38: 199-224.

Buch-Kromann, M., Gylling, M., Knudsen, L.J., Korzen, I & Müller, H.H. 2010. *The inventory of linguistic relations used in the Copenhagen Dependency Treebanks*. Technical report, Center for Research and Innovation in Translation and Translation

- Technology, Copenhagen Business School, <http://copenhagen-dependency-treebank.googlecode.com/svn/trunk/manual/cdt-manual.pdf>
- Carlson, L. & Marcu, D. 2001. *Discourse Tagging Reference Manual*. <http://www.isi.edu/~marcu/discourse/tagging-ref-manual.pdf>
- Carlson, L., Marcu, D. & Okurowski, M.E. 2001. Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. In *Proceedings of the 2nd SIGdial Workshop on Discourse and Dialogue*, Aalborg (Denmark) 1-2 September 2001.
- Hobbs, J.R. 1985. *On the coherence and structure of discourse*. Technical Report 85-37, Center for the Study of Language and Information (CSLI), Stanford University (CA).
- Korzen, I. 2003. Hierarchy vs. linearity. Some considerations on the relation between context and text with evidence from Italian and Danish. *Copenhagen Studies in Language* 29: 97-109.
- Korzen, I. 2004. Anafora associativa: aspetti lessicali, testuali e contestuali. In N. Maraschio & T. Poggi Salani (eds), *Italia linguistica anno Mille, Italia linguistica anno Duemila. Atti del XXXIV Congresso Internazionale di Studi della Società di Linguistica Italiana (SLI)*. Roma: Bulzoni, 593-607.
- Korzen, I. 2007. Mr. Bean e la linguistica testuale comparativa. Considerazioni tipologico-comparative sulle lingue romanze e germaniche. In M. Barbera, E. Corino & C. Onesti (eds), *Corpora e linguistica in rete*. Perugia: Guerra, 209-224.
- Korzen, I. 2009. Struttura testuale e anafora evolutiva: tipologia romanza e tipologia germanica. In I. Korzen & C. Lavinio (eds), *Lingue, culture e testi istituzionali*. Firenze: Franco Cesati, 33-60.
- Korzen, I. & Buch-Kromann, M. (in stampa). Anaphoric relations in the Copenhagen Dependency Treebanks. Deve apparire in *Proceedings of Beyond Semantics. DGfS Workshop*, Göttingen (Germany), 23-25 February 2011.
- Kromann, M.T. 2004a. The Danish Dependency Treebank and the DTAG treebank tool. In H. Holmboe (ed.), *Nordisk Sprogteknologi. Årbog for Nordisk Sprogteknologisk Forskningsprogram 2000-2004*. Copenhagen: Museum Tusulanums Forlag.
- Kromann, M.T. 2004b. *DTAG dependency annotation tool*. <http://isvcbs.dk/users/mbk/dtag/>
- Kromann, M.T. 2005. *DTAG howto*. <http://isvcbs.dk/~mbk/dtag/dtag-howtoT.html>
- Kromann, M.T., Lyng, S.K. & Mikkelsen, L. 2004. *Danish Dependency Treebank. Annotation Manual*. Department of Computational Linguistics, Copenhagen Business School. <http://www.buch-kromann.dk/matthias/treebank/>
- Lyons, J. 1977. *Semantics*. Cambridge: Cambridge University Press.
- Mann, W.C., Matthiessen, C. & Thompson, S.A. 1992. Rhetorical Structure Theory and Text Analysis. In W.C. Mann & S.A. Thompson (eds), *Discourse Description. Diverse Linguistic Analyses of a Fund-Raising text*. Amsterdam: Benjamins, 39-78.
- Mann, W.C. & Thompson, S.A. 1987. Rhetorical Structure Theory. A Theory of Text Organization. Information Sciences Institute, Los Angeles (CA), ISI/RS-87-190, 1-81.
- Matthiessen, C. & Thompson, S.A. 1988. The structure of discourse and 'subordination'. In J. Haiman & S.A. Thompson (eds), *Clause Combining in Grammar and Discourse*. Amsterdam: Benjamins, 275-329.
- Mladová, L., Zikánová, Š & Hajičová, E. 2008. From Sentence to Discourse: Building an Annotation Scheme for Discourse Based on Prague Dependency Treebank. In

- Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC 2008*, Marrakech (Morocco) 28-30 May 2008. Paris: ELRA.
- Montemagni, S., Barsotti, F., Battista, M., Calzolari, N., Corazzari, O., Lenci, A., Zampolli, A., Fanciulli, F., Massetani, M., Raffaelli, R., Basili, R., Pazienza, M.T., Saracino, D., Zanzotto, F., Mana, N., Pianesi, F., Del Monte, R. 2003a. The syntactic-semantic treebank of Italian. An overview. *Linguistica Computazionale XVI-XVII*: 461-492.
- Montemagni, S., Barsotti, F., Battista, M., Calzolari, N., Corazzari, O., Lenci, A., Zampolli, A., Fanciulli, F., Massetani, M., Raffaelli, R., Basili, R., Pazienza, M.T., Saracino, D., Zanzotto, F., Mana, N., Pianesi, F., Del Monte, R. 2003b. Building the Italian Syntactic-Semantic Treebank. In A. Abeillé (ed.), *Treebanks. Building and using Parsed Corpora*. Dordrecht: Kluwer, 189-210.
- Prasad, R., Dinesh, N., Lee, A., Joshi, A. & Webber, B. 2006. Annotating Attribution in the Penn Discourse TreeBank. In *Proceedings of the Workshop On Sentiment And Subjectivity In Text*, Sydney (Australia) 2006, Association for Computational Linguistics, 31-38.
- Prasad, R., Miltsakaki, E., Dinesh, N., Lee, A., Joshi, A., Robaldo, L. & Webber, B. 2007. *The Penn Discourse Treebank 2.0. Annotation Manual*. Technical report, University of Pennsylvania Institute for Research in Cognitive Science.
- Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi A. & Webber, B. 2008. The Penn Discourse TreeBank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC 2008*, Marrakech (Morocco) 28-30 May 2008. Paris: ELRA.
- Pustejovsky, J. 1995. *The Generative Lexicon*. Cambridge (MA): The MIT Press.
- Skytte, G. 1999a. Mr. Bean in danese e in italiano. Presentazione di una ricerca di linguistica testuale comparativa. In G. Skytte & F. Sabatini (eds), *Linguistica testuale comparativa*. Copenhagen: Museum Tusculanum Press, 295-303.
- Skytte, G. 1999b. Julekrybben/Il presepe. In G. Skytte, I. Korzen, P. Polito & E. Strudsholm (eds), *Strutturazione testuale in italiano e in danese*. Copenhagen: Museum Tusculanum Press, 419-483.
- Stede, M. 2008. Disambiguating Rhetorical Structure. *Research on Language and Computation* 6: 311-332.
- Taboada, M. & Mann, W.C. 2006a. Rhetorical Structure Theory: looking back and moving ahead. *Discourse Studies* 8(3): 423-459.
- Taboada, M. & Mann, W.C. 2006b. Applications of Rhetorical Structure Theory. *Discourse Studies* 8(4): 567-588.
- Webber, B. 2004. D-LTAG: extending lexicalized TAG to discourse. *Cognitive Science* 28: 751-779.
- Webber, B. 2006. Accounting for Discourse Relation: Constituency and Dependency. In M. Dalrymple (ed.), *Festschrift for Ron Kaplan*. Stanford (CA): CSLI Publications, 1-22.
- Wolf, F. & Gibson, E. 2005. Representing Discourse Coherence: A Corpus-Based Study. *Computational Linguistics* 31(2): 249-287.

SPANISH IN A TREEBANK PERSPECTIVE.  
THE COPENHAGEN DEPENDENCY TREEBANK PROJECT.  
MORPHOLOGICAL ASPECTS<sup>1</sup>

Henrik Høeg Müller  
Copenhagen Business School

**1. Introduction**

This article falls into two parts: one which is concerned with the morphological component of CDT, and another which outlines a hypothesis about NP-structure in Romance vs. Germanic languages, which the CDT can be used to prove or disprove.

**2. Morphology**

The morphological annotation in the CDT treebanks is only concerned with derivation and composition, since inflectional morphology can be identified and analysed automatically with a high degree of accuracy for all the languages involved in the treebanks. The internal structure of words and word-like phrases is encoded as a dependency tree which can be specified in two different ways: either as an ordinary dependency tree (the *dependency notation* in figure 2, left) or by means of an abstract representation of how the dependency tree for a morphologically complex word is constructed from roots annotated as lemmas combined with morphological operators (the *operator notation* in figure 2, right). In other words, the dependency notation to the left specifies the tree directly, whereas the operator notation to the right shows how the tree can be created from a set of operators. In the treebank annotation, we use the dependency notation to encode dependency structure between word-like phrases, i.e. phrasal compounds, while the operator notation is employed to encode dependency structure within solid orthography

---

<sup>1</sup> For a thorough presentation of the project and the principles behind the annotation, see e.g. Kromann (2003), Buch-Kromann (2006) and Buch-Kromann et al. (2007).

Cresti E., Korzen I. (eds), *Language, Cognition and Identity. Extensions of the endocentric / exocentric language typology* © 2010 Firenze University Press.

compounds and derivations. The CDT dependency notation in general analyses structures on the phrasal level, but marks all word-like phrases with a hash symbol (cfr. figure 1). In figure 1 the head appears in red and is marked as 1, and the non-head is black and marked as 0. The top arrow indicates that ‘shark’ is a noun object attached to the head as subjective argument, while the arrow underneath specifies the semantic relation.



Figure 1. Annotation of the word-like phrase ‘shark attack’ in CDT

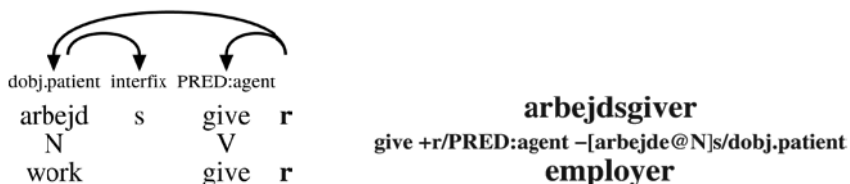


Figure 2. Morphological analysis annotated in dependency notation (left) and operator notation (right)

The dependency notation and the operator notation look very different, but they are merely two notational variants for the same underlying dependency tree. The operator notation maps on to a dependency structure with equivalent principles to the ones governing syntactic and discursive expansions.

In the operator notation, a morphological complex word is annotated as a lemma followed by one or more operators: ‘lemma op<sub>1</sub> op<sub>2</sub>...’. The operators are applied in order, and each operator is a functional-semantic instruction as to how the root, annotated as a lemma, or an already complex lemma, is to be interpreted when under the scope of a particular operator. The operator can both indicate a relation established by a traditional affix or the non-head of a compound. The operator specifies the position of the affix with respect to its base (cfr. prefix position-{}), infix position (‘~’), suffix position (‘+’) or combinations of these). Optionally, information about word class can be specified, indicating a non-standard word class of the stem or modifier.

The Spanish (1), English (2), and Danish (3)-(4) examples below illustrate how the morphological part of the annotation system works.



- |     |                                |   |
|-----|--------------------------------|---|
| (1) | antihéroe [antihero]           | héroe –anti/NEG:oppo                          |
| (2) | reannouncement                 | announce –re/ASPEC:iter +ment/PRED:core       |
| (3) | arbejdsgiver [employer]        | give +r/PRED:agent –[arbejde@N]s/dobj.patient |
| (4) | arbejderbolig [worker’s house] | bolig –[arbejd@V +er/PRED:agent]/FUNC         |

The minimally complex structure is exemplified in (1). Here the root ‘héroe’ is in scope of the operator ‘–anti/NEG:oppo’. The minus sign indicates that ‘anti-’ is a prefix. Moreover, it induces the meaning of negation/sub-category: opposition, and functions as dependent to the governing root. Generally the root is governor (head) and the element activating the morphological operation functions as dependent. However, when the operator is transformational, the operator functions as governing head and the root/stem as its dependent. This type of information is not written explicitly in the operator annotation.

In (2), apart from the prefix operator ‘–re/ASPEC:iter’, which shows that ‘re-’ is a prefix inducing an iterative aspectual meaning, the complex word consists of a suffix – annotated ‘+ment/PRED:core’ – which transforms the root ‘announce’ into a predicative eventive (core) noun. In terms of dependency structure, the governor is the suffix which takes the root as dependent element, and the prefix functions as dependent to the root.

The analysis of the compound in (3) can be explained as follows: The head is ‘giver’ [giver], which is a derivationally complex lexeme. The operator ‘+r/PRED:agent’ indicates that the head is an agent nominalisation of the verb ‘give’ [give] triggered by the suffix ‘-er’. The annotation of the non-head, i.e. ‘–[arbejde@N]s/dobj.patient’ indicates its pre-head position, that the lexical material is a noun with the interfix ‘-s’ (cfr. ‘[arbejde@N]s’) and that it corresponds to a direct object with the semantic function of Patient. The governor is the suffix which takes the root as dependent, and the non-head functions as dependent to the root.

In (4), the Danish word ‘arbejderbolig’ [worker’s house] is constructed from the root noun ‘bolig’ [house] by combining it with the complex stem ‘arbejder’ [worker] by means of a FUNCTIONAL composition relation.

The most important relation types in the morphological annotation are listed in table 1 (the examples are not always fully analysed, i.e., the real treebank annotation sometimes includes other derivational structures in addition to the exemplified relation). In the table, relation types with head-switching are italicised. The inventory of relations is inspired by Varela & Martín García (1999), Rainer (1999) and Müller (2001; 2003; 2006).

Table 1. The main relation types in the morphological annotation (relation types with head-switching are italicised)

<b><u>Relations that typically appear with prefixes</u></b>	
<b>LOC:pos</b>	(position: ‘intramural’ = <i>mural -intra/LOC:pos</i> )
<b>LOC:dir</b>	(direction/origin: ‘deverbal’ = <i>verbal -de/LOC:dir</i> )
<b>TIME:prec</b>	(temporal precedence: ‘prehistorical’ = <i>historical -pre/TIME:prec</i> )
<b>TIME:succ</b>	(temporal succession: ‘postmodernism’ = <i>modernism -post/TIME:succ</i> )
<b>NEG:oppo</b>	(opposition: ‘antihero’ = <i>hero -anti/NEG:oppo</i> )
<b>NEG:priv</b>	(privation: ‘desalt’ = <i>salt -de/NEG:priv</i> )
<b>GRAD:size</b>	(size/quantity: ‘minibar’ = <i>bar -mini/GRAD:size</i> )
<b>GRAD:qual</b>	(quality: ‘supercomputer’ = <i>computer -super/GRAD:qual</i> )
<b>ASPEC:rev</b>	(reversion: ‘deactivate’ = <i>activate -de/ASPEC:rev</i> )
<b>ASPEC:iter</b>	(iterative: ‘rebirth’ = <i>birth -re/ASPEC:iter</i> )
<b>ASPEC:cause</b>	(causative: ‘acallar’ [silence] = <i>callar -a/ASPEC:cause</i> )
<b>ASPEC:reflex</b>	(reflexive: ‘autopilot’ = <i>pilot -auto/ASPEC:reflex</i> )
<b>ASPEC:term</b>	(terminative: ‘oplåse’ [open] = <i>låse -op/ASPEC:term</i> )
<b>ASPEC:resul</b>	(resultative: ‘fastnagle’ [rivet] = <i>nagle -fast/ASPEC:resul</i> )
<b>MOD:quant</b>	(quantification: ‘multicultural’ = <i>cultural -multi/MOD:quant</i> )
<b>MOD:man</b>	(manner: ‘maleducado’ [badly educated] = <i>educado -mal/MOD:man</i> )
<b>MOD:qual</b>	(qualification: ‘paleochristian’ = <i>christian -paleo/MOD:qual</i> )
<b>TRANS</b>	(transitivising: ‘påsejle’ [colide]: <i>sejle -på/TRANS</i> )
<b><u>Relations that typically appear with suffixes</u></b>	
<b>AUG</b>	(augmentative: ‘perrazo’ [big dog] = <i>perro +azo/AUG</i> )
<b>DIM</b>	(diminutive: ‘viejecito’ [little old man] = <i>viejo +ecito/DIM</i> )
<b>PEJ</b>	(pejorative: ‘vinacho’ [bad vine] = <i>vino +acho/PEJ</i> )
<b>DER:nv</b>	(noun→verb derivation: ‘salar’ [to salt] = <i>sal +ar/DER:nv</i> )
<b>DER:av</b>	(adjective→verb derivation: ‘darken’ = <i>dark +en/DER:av</i> )
<b>DER:vv</b>	(verb→verb derivation: ‘adormecer’ [lull to sleep] = <i>dormir -+[a][ecer]/DER:vv</i> )
<b>PRED:agent</b>	(agent derivation: ‘singer’ = <i>sing +er/PRED:agent</i> )
<b>PRED:core</b>	(core derivation: ‘exploitation’ = <i>exploit@V +ation/PRED:core</i> )
<b>QUAL</b>	(deadjectival noun: ‘bitterness’ = <i>bitter +ness/QUAL</i> )
<b>NOPRED:agent</b>	(agent derivation: ‘miller’ = <i>mill +er/NOPRED:agent</i> )
<b>NOPRED:cont</b>	(container derivation: ‘azucarero’ [sugar bowl] = <i>azucar +ero/NOPRED:cont</i> )
<b>DEVERB:pas.poten</b>	(deverbal adjective: ‘transportable’ = <i>transport +able/DEVERB:pas.poten</i> )
<b>DENOM:rel.norm</b>	(denominal adjective): ‘presidential’ = <i>president +ial/DENOM:rel.norm</i> )
<b><u>Relations that typically appear with compounds</u></b>	
<b>CONST</b>	(constitutive: ‘træbord’ [wooden table] = <i>bord -træ/CONST</i> )
<b>AGENT</b>	(agent: ‘politikontrol’ [police control] = <i>kontrol -politi/AGENT</i> )
<b>ORIGIN</b>	(origin: ‘rørsukker’ [cane sugar] = <i>sukker -rør/ORIGIN</i> )
<b>FUNC</b>	(function: ‘krigsskib’ [war ship] = <i>skib -[krig]s/FUNC</i> )
<b>POS</b>	(position: ‘loftlampe’ [ceiling lamp] = <i>lampe -loft/POS</i> )
<b>TIME</b>	(time: ‘oktoberregn’ [October rain] = <i>regn -oktober/TIME</i> )
<b>ABOUT</b>	(theme: ‘skattelov’ [tax law] = <i>lov -[skat]te/ABOUT</i> )

### 3. NP-hypothesis

It has often been pointed out as a systematic difference between Germanic and Romance languages that Germanic languages frequently use compounding to express what Romance languages convey by a derivational strategy (Bally 1932; Rainer & Varela 1992). A consequence of this is that many simple and derived words in e.g. Spanish have compounds as their translational equivalents in e.g. Danish (cfr. (5) and (6) below).

#### Simple noun → compound

- |      |                            |              |
|------|----------------------------|--------------|
| (5a) | berberecho – hjertemusling | [cockle]     |
| (5b) | búho – hornugle            | [horned owl] |
| (5c) | pulpito – prædikestol      | [pulpit]     |

#### Derivation → compound

- |      |                                      |   |
|------|--------------------------------------|---|
| (6a) | escritorio – skrivebord              | [writing desk] (escribir: write)          |
| (6b) | dentadura – tandsæt                  | [set of teeth] (diente: [tooth])          |
| (6c) | petrolero – olietankskib/oliehandler | [oil dealer/oil tanker] (petróleo: [oil]) |

It is here assumed that this cross-linguistic contrast can be accounted for by means of a lexical-typological hypothesis which builds on the distinction between *endo- and exocentric languages*.

#### 3.1 Exocentric vs. endocentric languages

The following short presentation of the differences between exocentric and endocentric languages takes its point of departure in the typological basic assumption that the Romance languages concentrate information in the nominal arguments of the verb, while the verb itself is left relatively underspecified (exocentric = concentration of information in the periphery of the sentence), which prototypically results in a description of the relevant state of affairs as abstract relations between specific entities. Complementary to this distribution of information, the Germanic languages concentrate information in the verb and leave the nominal argument relatively underspecified (endocentric = concentration of information in the centre of the sentence), which prototypically results in a description of the relevant state of affairs as rather concrete relations between underspecified entities (for a more detailed description see e.g. Herslund & Baron 2005).

3.1.1 *Verbs*

The postulated informational or lexical specificity of the Danish endocentric verbs is based on the insight that these verbs generally lexicalise the semantic component MANNER and, as a result of that, impose selectional restrictions mainly on the subjective argument with which they combine. As illustrated in the left hand column of (7), the Danish expressions corresponding to the Spanish verb ‘*entrar*’ (enter) are composed by a verb containing the MANNER component and a particle of directionality ‘*ind*’ (into), and they adjust semantically to the usual manner of “entering” performed by the subject referents. On the contrary, the Spanish verb ‘*entrar*’ (enter) does not contain any information on the manner in which the motion is carried out, and consequently it combines with any subject whose referent can perform a movement, as shown in the right column of (7). In Spanish it is of course also possible to specify the manner in which the actions are realised. This can be done via adverbial satellites in the form of prepositional phrases or gerunds or through imperfective, unergative MANNER verbs such as ‘*bailar*’ (dance) or ‘*nadar*’ (swim). However, usually the MANNER component is only expressed in Spanish if this aspect of the verbal action is of specific importance or cannot be extrapolated from the situation. In contrast to this, the MANNER component is part of the lexical make-up of most Danish verbs – with some exceptions of course – and therefore Danish verbs are obliged to express the manner, they simply cannot avoid it.

(7)	Endocentric languages [Danish]		Exocentric languages [Spanish]		
	<i>hunden</i>	<b><i>løb ind</i></b>	<i>el perro</i>	<b><i>entró</i></b>	[the dog]
	<i>fisken</i>	<b><i>svømmede ind</i></b>	<i>el pez</i>	<b><i>entró</i></b>	[the fish]
	<i>fuglen</i>	<b><i>fløj ind</i></b>	<i>el pájaro</i>	<b><i>entró</i></b>	[the bird]
	<i>bilen</i>	<b><i>kørte ind</i></b>	<i>el coche</i>	<b><i>entró</i></b>	[the car]
	<i>skibet</i>	<b><i>sejlede ind</i></b>	<i>el barco</i>	<b><i>entró</i></b>	[the ship]
	<i>manden</i>	<b><i>gik ind</i></b>	<i>el hombre</i>	<b><i>entró</i></b>	[the man]

3.1.2 *Nouns*

As illustrated in (8) the nouns of the two language types show opposite lexicalisation structures to those of the verbs. Exocentric, mainly artefact denoting, nouns are generally lexically more contentful and precise than the endocentric ones. Whereas exocentric denominations of artefacts tend to lexicalise the semantic component FIGURE, i.e. the shape, dimensionality and structure of the object, endocentric artefact denoting nouns are inclined to lexicalise only the component FUNCTION, which is an inherent abstract feature of any artefact denoting noun. This exclusive focus on the purpose of the object, or non-focus on its form, means that Danish simple nouns in many cases are semantically vague and therefore they may function as denominations on a generic prototype level, i.e. a general hyperonymic

level, which for the above mentioned reason of lexicalisation of form in the Romance languages does not exist in Spanish. When in Danish a level below the general hyperonymic one is needed, this is usually achieved by means of nominal compounds as shown in (8). The nouns ‘vogn’ (wagon), an object used for transportation, and ‘tæppe’ (carpet), an object used for decoration or covering things, represent lexicalisation on the family level, while the Romance nouns, because of their lexical specification or content of outer appearance, must denote subtypes, i.e. object on a hyponymic level (for Italian see e.g. Korzen 2008). A predictable consequence of this difference is of course that the Danish nouns ‘vogn’ and ‘tæppe’ cannot be adequately translated into Romance without a proper context.

(8) Endocentric languages [Danish]	Exocentric languages [Spanish]	
<i>tæppe</i>	[Ø]	lit. transl.
<i>sengetæppe</i>	<i>colcha</i>	[bed –]
<i>vægtæppe</i>	<i>tapiz</i>	[wall –]
<i>ægte tæppe</i>	<i>alfombra</i>	[genuine –]
<i>væg-til-væg tæppe</i>	<i>moqueta</i>	[wall to wall –]
<i>teatertæppe/ scenetæppe</i>	<i>etelón</i>	[theatre/ stage –]
<i>slumretæppe</i>	<i>manta</i>	[slumber –]
<i>vogn</i>	[Ø]	lit. transl.
<i>personvogn</i>	<i>coche</i>	[person –]
<i>lastvogn</i>	<i>camión</i>	[load –]
<i>godsvogn</i>	<i>vagón</i>	[goods –]
<i>sækkevogn</i>	<i>carretilla</i>	[sack –]
<i>hyrevogn</i>	<i>taxi</i>	[hire –]
<i>ladvogn</i>	<i>plataforma</i>	[platform –]
<i>varevogn</i>	<i>furgoneta</i>	[goods –]

The shell fish taxonomy in (9) illustrates the same phenomenon. Spanish does not have a general hyperonym, and its hyponyms are not necessarily compounds as in Danish, where the hyponym level is predominantly lexicalised on the basis of composite nouns (see Herslund 1997: 31).

(9)	1. generic hyperonym, class	<i>(skaldyr – marisco)</i> [shell fish]
	2. general hyperonym, family	<i>(musling – Ø)</i>
	3. hyponyms, species	<i>(blåmusling, hjertemusling, kammusling, venusmusling – mejillón, berbereco, vieira, almeja)</i> [mussel, cockle, scallop, clam]

### 3.1.3 Perspectives

The two major word classes, nouns and verbs, are semantically complementary both within the single languages and across language types. Lexical information is organised differently in the Romance and Germanic languages due to the different distribution of semantic content on the two word classes.

The question is in this context, – how do these typological differences affect the issue of compounding?

A consequence of the semantic vagueness of the Danish simple nouns and, accordingly, their lexicalisation on a hyperonymic family level is that in order to designate entities on a hyponymic level Danish must make use of the compounding system. Conceptualising or communicating about concrete species of e.g. shell fish, wagons, etc. requires a word formation system that is capable of specifying the constitution, form, purpose or origin of the object in question and hence create an expression which denotes a subtype. Because of the great demand of such a mechanism, composition is very frequent in Danish and therefore incorporated into the grammatical system as a highly automated morphological word-formation process.

On the contrary, Romance simple nouns are already saturated in a semantic sense, so the Romance languages do not necessarily need, and consequently they have not developed, a full morphological system to deal with this information packaging task. Either the semantic components are already encapsulated in the simple noun or they use an alternative strategy, namely derivation, to lead the original lexical unit in another semantic direction. So although phrasal composition of the [N prep. N]-type also in the Romance languages is very often a prerequisite for creating subtype denoting lexical expression, it can be regarded as additional to the derivational system and the semantically contentful nouns. Therefore composition in the Romance languages is not routinised as part of a morphological system, but has the status of a syntactic device of a complementary nature.

Another reflection that supports the view of a fundamental typological difference between the two language types, and which influences their word-formation systems, is the fact that when the semantically underspecified Danish nouns – such as for example the lexeme ‘vogn’ in (8), which denotes some sort of mobile device for transportation with wheels – occur as heads in compounds, they resemble the semantically underspecified derivational suffixes of the Romance languages. This point is illustrated with the examples in (10).

(10)	<b>Spanish</b>	<b>Danish</b>	
	<i>punte</i> – <i>pontaje</i>	<i>bro</i> – <i>bropenge</i>	[bridge – bridge toll]
	<i>olmo</i> – <i>olmeda</i>	<i>elm</i> – <i>elmelund</i>	[elm – elm grove]
	<i>diente</i> – <i>dentadura</i>	<i>tand</i> – <i>tandsæt</i>	[tooth – set of teeth]
	<i>petróleo</i> – <i>petrolero</i>	<i>olie</i> – <i>oliehandler</i>	[oil – oil dealer]

<i>escribir – escritorio</i>	<i>skrive – skrivebord</i>	[write – writing desk]
<i>campana – campanario</i>	<i>klokke – klokketårn</i>	[bell – belfry]
<i>plátano – platanal</i>	<i>banan – bananplantage</i>	[banana – banana grove]

The analogy between the categories consists in the fact that both the Danish nouns and the Spanish suffixes are semantically vague in the sense that in isolation they do not convey any specific meaning. Both types of elements need to be lexically saturated from outside to obtain full “denotational status”. They share the feature of semantic unspecificity, but syntactically the Danish nouns function as heads and the Spanish suffixes as modifiers in relation to the nominal elements with which they unite. The analogy is further corroborated by the fact that the Danish nouns, when functioning as heads, are even reduced prosodically and pronounced with secondary stress.

This data substantiates the overall assumption that Danish nouns are semantically imprecise or ambiguous, and that Danish, as a result of that, has developed a nominal composition system which is a derivational morphological process invisible to syntax. The Romance languages have more difficulties in combining semantic heavy weight nouns, and for that reason they make comparatively more use of the derivational system or, of course, the phrasal compounding system which is predominantly a syntactic process.

Of course, a number of questions can be raised with respect to this hypothesis, but here we shall only mention a few of the most obvious ones:

- (i) The general characterisation of Danish nouns as being semantically vague is based on a specific subset of nouns, namely nouns denoting artifacts and which are typically equivalent to Spanish simple nouns or derivations. In this connection we must not forget that in many cases there is actually a structural parallelism between Spanish and Danish (cfr. e.g. ‘barco de guerra/ krigsskib’ [war ship], ‘granada de mano/ håndgranat’ [hand grenade] and ‘máquina de escribir/ skrivemaskine’ [type writer]) where both languages use compositional structures, contrary to what is indicated in (5a-c) and (6a-c) above.
- (ii) Also, it is not immediately evident that the generalisation made about Danish compounding (or Germanic in general, for that matter) is entirely valid, as heads of Danish compounds cannot immediately be judged semantically vague in all cases. This is especially evident when the head noun instead of an artifact denotes an abstract entity as in ‘indkomstskat/Einkommensteuer’ (income tax) or ‘regeringskonference/Regierungskonferenz’ (intergovernmental conference). So the question is how far this generalisation can be justified and in what sort of situations it will prove incorrect.

- (iii) In many cases the status of Romance syntagmatic structures is unclear as to whether they should be regarded as compound-like constructions or free syntactic phrase formations. In the Romance languages the compoundedness of phrasal structures is commonly measured by degree of lexicalization, i.e. a semantic concept, and not formal evidence as in the non-English Germanic languages. As the criteria for compounding vary from language type to language type, it is questionable on what basis the comparison is being made.

On this occasion the hypothesis has been introduced and substantiated with a limited number of examples, but there is no doubt that it is an empirical matter requiring more detailed studies and above all statistically significant data.

#### 4. Conclusion

In combination with the CDT parallel text alignment system<sup>2</sup>, which falls outside the scope of this article, the segmentation and analysis of morphological structure done in CDT provide a valuable tool for studying strategies of word and phrase formation across languages and language types. Hypotheses like the one presented above concerning morphological correlations between types of languages can be substantiated or rejected on solid statistical ground, and the CDT system can make available qualitative data that can lead to the refinement of already existing assumptions and, generally, to new insights into morphological monolingual issues and cross-linguistic contrasts. For example, in order to address the specific problem of Romance phrase formation vs. compounding, the annotators mark all word-like phrases with a hash symbol (cfr. footnote 2) which in correlation with the aligned decomposed Germanic compounds, derivation or other parallel structures, will provide important information for “solving the puzzle” of Romance phrase formation by means of free morphemes. Also, to reverse the aspect, the semantic relations between head and non-head in Germanic compounds can be correlated with corresponding structures in Romance languages (Spanish and Italian), in the form of either simple nouns, derivations or compounds, and in this way it will be possible to establish on an empirical foundation how the languages choose to encode information in different lexical, morphological or syntactic configurations. Moreover, the treebanks permit us to investigate a broad range of intralinguistic issues, e.g. whether or not a language has particular types of compounds or derivations, and how frequent this type of morphological structure is, which are the

---

<sup>2</sup> By means of a word alignment system the CDT is able to specify translational equivalents between source and target texts, i.e. the minimal word groups that correspond to each other with respect to meaning or function (see e.g. Buch-Kromann et al. 2009: 215).



preferred strategies of encoding different semantic relations (Agentive, Functional, Positional, etc.) in a given language, etc. Another important aspect is that the treebanks could be used as a source for automated tasks such as creating parsers and giving morphological input to machine translation systems.

## Bibliography

- Bally, C. 1932. *Linguistique générale et linguistique française*. Berne: Francke.
- Buch-Kromann, M. 2006. *Discontinuous Grammar. A dependency-based model of human parsing and language learning*. PhD diss., Copenhagen Business School.
- Buch-Kromann, M., Wedekind, J. & Elming, J. 2007. The Copenhagen Danish-English Dependency Treebank v. 2.0. <http://buch-kromann.dk/Matthias/cdt2.0>.
- Buch-Kromann, M., Korzen, I. & Müller, H.H. 2009. Uncovering the ‘lost’ structure of translations with parallel treebanks. *Copenhagen Studies in Language* 38: 199-224.
- Herslund, M. (ed.). 1997. *Det franske sprog. Kapitel I. Grundlag*. Preliminary version. Copenhagen: Copenhagen Business School.
- Herslund, M. & Baron, I. 2005. Langues endocentriques et langues exocentriques. Approche typologique du danois, du français et de l’anglais. *Langue française* 145 : 35–53.
- Korzen, I. 2008. Determination in endocentric and exocentric languages. With evidence primarily from Danish and Italian. In H.H. Müller & A. Klinge, *Essays on Nominal Determination. From Morphology to Discourse Management*. Amsterdam: Benjamins.
- Kromann, M. T. 2003. The Danish Dependency Treebank and the DTAG treebank tool. In *Proceedings of the Second Workshop on Treebanks and Linguistic Theories, TLT 2003*, 14-15 November, Växjö, 217–220.
- Müller, H.H. 2001. Spanish N de N-structures from a cognitive perspective. In I. Baron, M. Herslund & F. Sørensen (eds), *Dimensions of Possession*. Amsterdam: Benjamins, 169-186.
- Müller, H.H. 2003. Strategies de lexicalisation des noms composés en espagnol. In M. Herslund (ed.), *Aspects linguistiques de la traduction*. Bordeaux: Presses Universitaires de Bordeaux, 55-84.
- Müller, H.H. 2006. *Nominalkomposition i moderne spansk. En teori om betydningsdannelse*. PhD diss., Copenhagen Business School.
- Rainer, F. 1999. La derivación adjectival. In I. Bosque & V. Demonte (eds), *Gramática Descriptiva de la Lengua Española*, volume 3. Madrid: Espasa, 4595–4643.
- Rainer, F. & Varela, S. 1992. Compounding in Spanish. *Rivista di Linguistica* 4(1): 117–142.
- Varela, S. & Martín García, J. 1999. La prefijación. In I. Bosque & V. Demonte (eds), *Gramática Descriptiva de la Lengua Española*, volume 3. Madrid: Espasa, 4993–5040.



# LE RISORSE DI RETE DELL'ITALIANO. PRESENTAZIONE DEL PROGETTO “RIDIRE.IT”

Massimo Moneglia, Samuele Paladini

LABLITA, Università di Firenze

## 1. La crescita della rete e le lingue

L'italiano è attualmente parlato da circa 75 milioni di persone e si colloca tra la quindicesima e la ventesima posizione tra le lingue più parlate al mondo<sup>1</sup>. In Europa è una delle 23 lingue ufficiali dell'Unione. L'italiano gode inoltre di una notevole possibilità di intercomprensione all'interno del gruppo romanzo<sup>2</sup>, che conta quasi 900 milioni di parlanti ed è quindi, nell'insieme, tra le entità linguistiche più rappresentative, accanto all'inglese, il mandarino e l'indi-urdu.

Le lingue, la loro importanza e le politiche di diffusione e salvaguardia ad esse dedicate debbono però essere rapportate non solo ai parlanti attuali, ma al contesto della comunicazione globale nel quale attualmente esse si trovano a vivere e al rilievo culturale e pratico che in tale contesto esse rivestono. Internet è attualmente il più grande deposito di informazione linguistica esistente ed è insieme ambiente e mezzo privilegiato dell'uso di una lingua, lo spazio entro il quale sia gli usi funzionali sia gli usi creativi del linguaggio sono esercitati con sempre maggior frequenza.

Lo sviluppo di Internet dà la misura di questa asserzione. L'indice di Google contava nel 1998 26 milioni di pagine, un miliardo alla fine del 2000 e si calcola che attualmente le pagine web indicizzate dai motori di ricerca siano nell'ordine delle decine di miliardi.

La figura 1 riporta la crescita dei domini registrati e dei domini attivi, che, come si può vedere, si impenna dalla metà di questa decade mostrando un ritmo di crescita geometrico<sup>3</sup>.

---

<sup>1</sup> Cfr. Linguasphere.

<sup>2</sup> Cfr. Blanche-Benveniste et al. (1997).

<sup>3</sup> Le statistiche sul web riportate in figura 1, figura 2 e figura 4 sono tratte da Internet World Stat - Usage and Population.

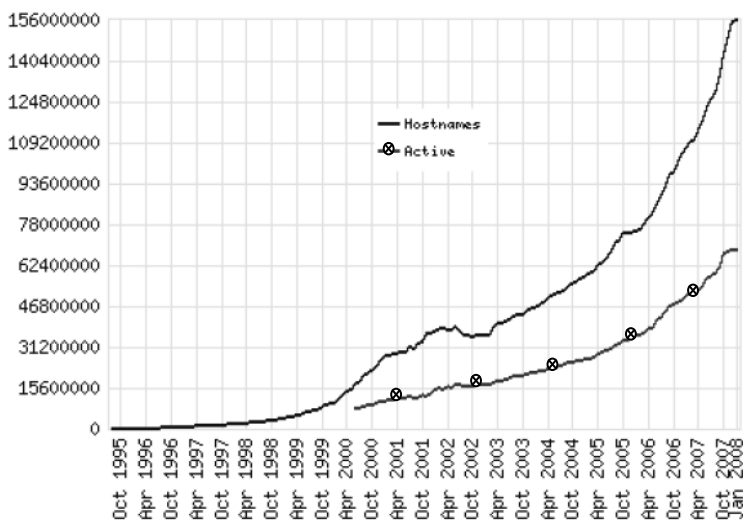


Figura 1. Totale dei domini registrati e attivi (periodo: agosto 1995 - gennaio 2008)

Lo sviluppo del web può essere monitorato attraverso una prospettiva geografica che individua la provenienza degli utenti e indirettamente definisce il potenziale di crescita della rete stessa nelle varie aree del mondo.

Le statistiche in figura 2 mostrano che, nonostante la crescita sia stata geometrica, la possibilità dello sviluppo globale della rete dal punto di vista degli utenti è enorme, in quanto riguarda più del 70% della popolazione. Approssimativamente, su una popolazione mondiale di circa sei miliardi e ottocentomila persone, gli utenti di Internet sono attualmente circa due miliardi. Le statistiche a disposizione permettono di osservare le differenze di penetrazione della rete nelle diverse aree del mondo e anche di osservare come sta evolvendo la crescita di questo media.

Da questo punto di vista, se il nord America è sostanzialmente saturo, in Europa esiste ancora un margine di crescita che sta però rapidamente portando anche il vecchio continente ad una sostanziale copertura della maggior parte della popolazione. Il maggiore potenziale di crescita si registra in Africa, e poi rispettivamente in Asia, Medio oriente ed America latina, aree nelle quali l'accesso alla rete è ancora basso rispetto alla popolazione, ma nelle quali, in questa fase, il ritmo di espansione sfiora il 5% della popolazione globale per anno.

In altri termini, nei poco meno di due anni in considerazione, gli utenti sono aumentati di circa 500.000 unità, ovvero il 25% degli utenti attuali sono nuovi utenti.

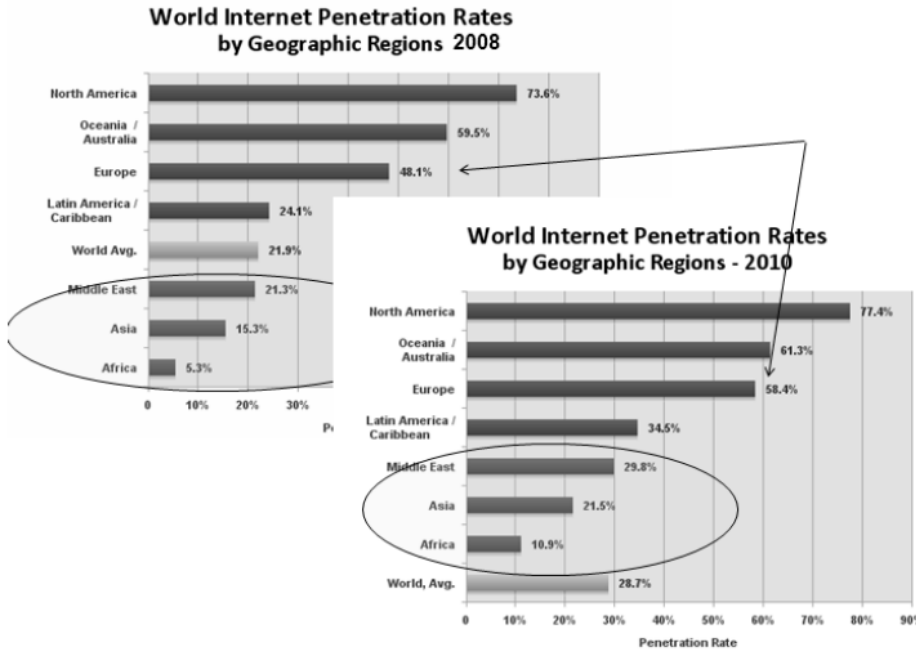


Figura 2. La penetrazione di Internet nelle diverse regioni del mondo nel 2008 e nel 2010

Considerare le prospettive di una lingua nel mondo globale rende quindi necessario assumere come prospettiva privilegiata la possibilità che questa ha di vivere ed esprimersi su questo canale, nel quale il problema linguistico è capitale.

Alla fine degli anni '90 si poteva ancora dire che l'inglese era la lingua del web (cfr. figura 3). Come si può vedere dalla tabella 1, questo dato si è rapidamente modificato: già nel 2002 circa la metà del web era in lingue diverse dall'inglese e grosso modo, in una percentuale simile, lingue diverse dall'inglese erano usate dagli utenti per accedere ai servizi della rete.

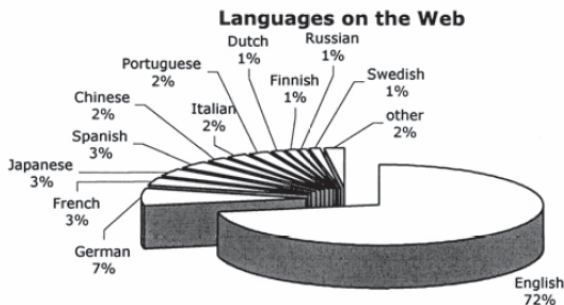


Figura 3. Proporzion delle lingue nel web da un campione casuale di pagine (1999)<sup>4</sup>

<sup>4</sup> Figura tratta da O'Neill, Lavoie & Bennett (2003).

Nel 2005 l'inglese era però già sceso al 30% e viene stimato nel 2008 da Internet World Stat intorno al 20% (figura 4).

Tabella 1<sup>5</sup>

<b>Internet Statistics: Distribution of languages on the Internet 2002 (millions of web pages by language)</b>			<b>Languages used to access Google</b>		
<b>Total web pages</b>	2024,7			2002	2001
English	1142,5	56,40%	English	57%	-64%
German	156,2	7,70%	German	12%	-9%
French	113,1	5,60%	Japanese	7%	-8%
Japanese	98,3	4,90%	Spanish	6%	-5%
Spanish	59,9	3,00%	French	5%	-4%
Chinese	48,2	2,40%	Chinese	3%	-1%
Italian	41,1	2,00%	Italian	2%	-2%
Dutch	38,8	1,90%	Other	8%	-4%
Russian	33,7	1,70%			
Korean	30,8	1,50%			
Portuguese	29,4	1,50%			
Swedish	15,1	0,70%			
Polish	14,8	0,70%			
Danish	12,3	0,60%			
Czech	11,5	0,60%			
Turkish	4,9	0,20%			
Hungarian	4,1	0,20%			
Greek	2	0,10%			
Other	168	8,30%			

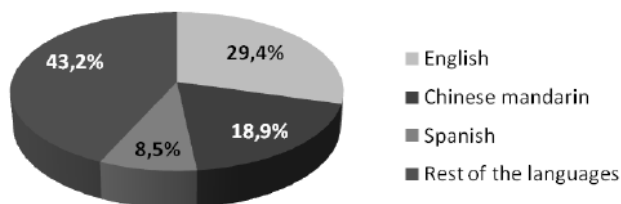


Figura 4. Le lingue di Internet nel 2008

<sup>5</sup> Informazione tratta da Gerrand (2007).

Il numero e l'origine geografica degli utenti della rete è un indice indiretto delle lingue degli utenti (figura 5) che sono calcolati, come si diceva, in circa un miliardo e mezzo al 2008 e in circa 2 miliardi attualmente. Asia, Europa e Nord America costituiscono in quel rilevamento la gran parte di questa cifra e l'inglese e il cinese mandarino coprono da soli circa la metà dell'utenza. Significativamente, l'italiano è presente nella top ten delle lingue degli utenti di Internet nel 2008 (in decima posizione), ma ne esce nel 2010:

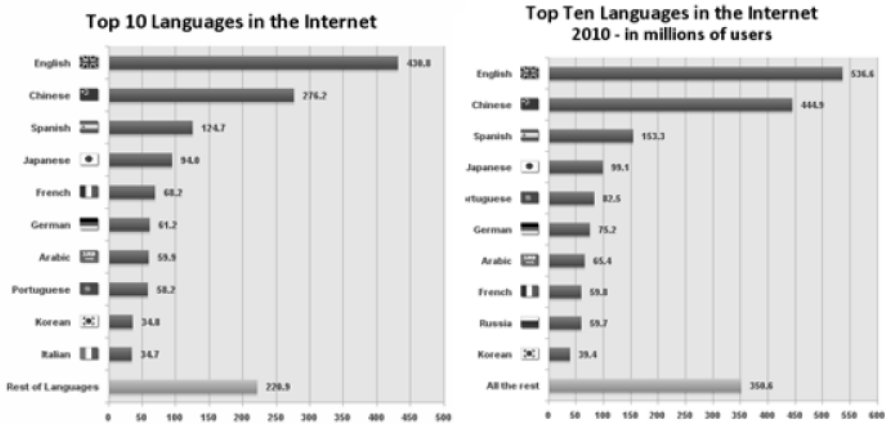


Figura 5. Raffronto 2008-2010 del numero di utenti di Internet per area geografica

Nel confronto relativo al breve periodo si evidenzia da un lato l'impressionante aumento di utenti cinesi, dall'altro l'aumento della diversificazione di origine geografica (e quindi linguistica) degli utenti del web (*All the rest*). Parallelamente le lingue che hanno un numero elevato di parlanti si affacciano sempre più massicciamente sulla rete (è il caso del portoghese e del russo).

In sintesi, il web è multilingue sia dal punto di vista dei contenuti sia dal punto di vista degli utenti. Inglese, Spagnolo e Mandarino coprono da soli il 60% della rete, e, si può dire, sono le attuali lingue di Internet, ma il restante 40%, che tenderà sempre più crescere, parla le molte lingue degli utenti.

In altri termini, l'inglese è la lingua franca nel mondo globale, ma questo fatto non ha cancellato la diversità linguistica che esiste e si esprime nella rete rispondendo alle necessità funzionali e culturali dei suoi utenti. Potremmo sintetizzare il rapporto tra crescita del web e sua espressione linguistica notando una correlazione inversa tra l'espansione della rete e la dominanza dell'inglese come lingua del web.

## 2. L'italiano nella Rete

Via via che si affacciano alla rete le popolazioni di tutte le aree del mondo, l'italiano necessariamente esce dalla top ten delle lingue del web. Tuttavia, la lingua italiana è già ampiamente rappresentata in rete, in quanto il web attira già più di 30 milioni di utenti, raddoppiati nell'arco di 10 anni, con un indice di penetrazione importante, sebbene più basso rispetto alla media europea: 51,7 % rispetto al 58,4%.

Il ranking dell'italiano in questo quadro ha dunque prospettive di crescita appena superiori alle altre lingue di cultura europee (francese e tedesco in particolare) ma, all'interno dell'incremento globale, la sua incidenza sull'insieme non può che diminuire. Diversa è la prospettiva dello spagnolo e anche del portoghese, che sia per la numerosità della popolazione sia per le larghe possibilità di crescita di Internet nelle aree di competenza di queste lingue possono aumentare considerevolmente il loro impatto.

La prospettiva della lingua italiana nel contesto della rete è strettamente legata a due fattori di ordine diverso che non hanno diretto rapporto con la numerosità della popolazione e l'entità del suo accesso alla rete. Da un lato l'interesse globale per i contenuti italiani e per l'universo culturale rappresentato attraverso la lingua italiana, dall'altro gli strumenti che consentono l'accesso a tali contenuti, la loro diffusione e la loro localizzazione in contesti linguistici e culturali diversi.

La globalizzazione ha una tendenza alla omologazione. Nonostante la cultura italiana abbia largo impatto nel mondo (per esempio si stima che attualmente negli Stati Uniti l'italiano è la quarta lingua più insegnata come L2 dopo lo spagnolo, l'arabo e il cinese), deve essere chiaro che i contenuti italiani non sono, né in prospettiva saranno, dominanti a livello globale. Questo dato può essere valutato semplicemente attraverso gli strumenti della stessa rete, osservando come sono usate le lingue per fare ricerche nel web e quali contenuti interessano l'utenza.

Google ha recentemente pubblicato, in un suo progetto ambiziosamente chiamato "Spirito del tempo" (ZeitGeist), statistiche relative alle parole più utilizzate per fare ricerche sia a livello globale che nei bacini nazionali corrispondenti alle varie lingue. La statistica fornisce quindi un indice indiretto dell'impatto di una lingua e di un universo culturale sulla rete stessa.

Gli ordinamenti seguenti identificano le parole digitate dagli utenti che hanno avuto rapidi picchi di crescita nel 2008, rispettivamente a livello globale e a livello italiano. Le *keyword* ricercate identificano i contenuti sui quali l'attenzione degli utenti di Internet si è focalizzata in modo massiccio e concentrato e, grossolanamente, rappresentano l'universo di riferimento ai livelli globale e locale.



Tabella 2. Ricerche che hanno avuto picchi nel 2008 a livello globale e italiano (da ZeitGeist)

Ricerche in ascesa (globale)	Ricerche in ascesa (italiano)
1. sarah palin	1. pechino 2008
2. beijing 2008	2. facebook
3. facebook login	3. obama
4. tuenti	4. ecopass
5. heart ledger	5. la talpa
6. obama	6. finanziaria 2008
7. nasza klasa	7. saviano
8. wer kennt wen	8. wiki
9. euro 2008	9. mutui
10. jonas brothers	10. cinquecento

A livello globale, non emerge nessun riferimento italiano ad altissimo impatto, mentre è presente l'ambito più generalmente europeo (europei di calcio, *social network* europei).

Al contrario, le *keyword* con maggiore impatto a livello globale sono anche appannaggio degli utenti italiani (pechino 2008, facebook, obama), che manifestano però, per quanto riguarda i riferimenti cercati in rete, un saldo ancoraggio alla specificità del contesto economico e culturale italiano (la talpa, finanziaria 2008, saviano, mutui, cinquecento)<sup>6</sup>.

Questa grossolana osservazione consente semplici deduzioni. La prospettiva più concreta di sviluppo nel web di una lingua come l'italiano è da un lato nell'utilizzo domestico e dall'altro nel rafforzamento dell'impatto dei suoi contenuti negli ambiti di eccellenza sia culturale sia funzionale, contenuti che certo non possono aspirare oggi ad assumere un ruolo dominante nell'universo globale.

### 3. Il progetto RIDIRE: un web corpus per l'acquisizione dell'italiano

#### 3.1 Il progetto RIDIRE

Per garantire il ruolo globale di una lingua è necessario rafforzarne la possibilità di fruizione e utilizzo dei suoi contenuti nel web nel contesto globale multilingue. Ciò

<sup>6</sup> A livello globale, espressioni appartenenti alla sfera culturale italiana emergono dai dati pubblicati da Google solo se si considerano i disaggregati relativi a domini specifici dai quali si vince che l'italiano entra nelle top ten delle *keyword* usate solo per le ricerche nel dominio della cucina e degli aperitivi!

può avvenire sia attraverso la traduzione nella lingua dell'utente (localizzazione) sia, più comunemente, attraverso l'utilizzo della lingua veicolare, ma è evidente che senza un interesse per l'accesso diretto ai contenuti il valore funzionale e culturale della lingua tende progressivamente a diminuire.

Per l'italiano si debbono valutare due condizioni specifiche: (a) l'italiano ha sia un patrimonio culturale attuale sia un'eredità culturale di valore globale; (b) l'italiano è disseminato nel mondo attraverso una vasta presenza di comunità semi-italofone, conseguenza dei fenomeni migratori del XIX e XX secolo. Inoltre l'italiano è una delle scelte più frequenti tra gli apprendenti una seconda lingua nel mondo, proprio per l'impatto del suo patrimonio culturale.

Il progetto RIDIRE.it (Risorsa Dinamica Italiana di Rete), promosso dalla Società Internazionale di Linguistica e Filologia Italiana (SILFI) in collaborazione con un consorzio di università italiane e finanziato sul Fondo Italiano per la Ricerca di Base (FIRB), si occupa del rapporto tra potenziamento della lingua e accesso ai contenuti italiani della rete<sup>7</sup>.

Se l'accesso e la fruizione dei contenuti italiani rappresentati sulla rete è una via essenziale per il contatto con la cultura italiana nella società dell'informazione, ciò necessita il rafforzamento della conoscenza della lingua italiana, specificamente in coloro che per motivi di studio, di lavoro o identitari sviluppano un interesse per tali contenuti. Il Progetto RIDIRE.it è rivolto agli italiani all'estero di seconda e terza generazione e a chi studia l'italiano come L2, con particolare riguardo al contesto della formazione universitaria, ed è dedicato a formare un web corpus composto dai contenuti della rete più rappresentativi per la cultura italiana. La risorsa è concepita come una infrastruttura di servizi linguistici in rete che da un lato danno accesso ai contenuti propri della vita italiana culturale e pratica, e al contempo sfruttano le potenzialità e la fraseologia insite in quegli stessi contenuti per il consolidamento del possesso della lingua italiana.

## 3.2 La rappresentatività dei web corpora e i corpora italiani in rete

### 3.2.1 Il problema se i web corpora, o comunque l'informazione linguistica complessivamente costituita dalla rete, fossero o meno una sorgente adeguata a rappresentare le lingue si è posto esplicitamente all'inizio degli anni 2000, in coincidenza con la prima espansione di Internet (Ide et al. 2002; Kilgariff 2001). L'opportunità offerta ai linguisti dall'emergere di una gran mole di dati linguistici in formato computabile (al tempo, come abbiamo visto, per la gran parte in lingua

---

<sup>7</sup> Coordinamento: Società internazionale di linguistica e filologia italiana (SILFI). Partners: UNIFI (LABLITA); UNIFI (Dipartimento Sistemi e Informatica), UNITO (Dipartimento. Scienze Letterarie e Filologiche); UNIROMA3 (Dipartimento di Italianistica); UNINA (Dipartimento di Filologia Moderna); UNISI (Dipartimento di Economia).

inglese) rappresentava infatti un terreno estremamente promettente per la linguistica dei corpora e la linguistica computazionale.

Il modello di rappresentazione dell'uso linguistico attraverso corpora più accreditato, fornito dal *British National Corpus* e concepito nei primi anni '90 (Atkins et al. 1992; Burnard 2000), prevede però che i corpora abbiano al loro interno una variazione complessa per genere e funzione dei testi e una pesatura delle varie componenti all'interno del corpus che approssima la diversa rappresentatività dei generi e il loro diverso impatto nell'universo di riferimento, oltre che i diversi canali attraverso i quali l'informazione linguistica è diffusa. Questo modello pone problemi sostanziosi quando si applica alla rete. Questa si configura infatti come un insieme amorfo di dati linguistici, che posseggono cifre linguistiche specifiche, proprie della comunicazione in rete.

In estrema sintesi, l'informazione linguistica in rete è da un lato opaca per quanto concerne il peso delle sue componenti sull'insieme e dall'altro, ancor più criticamente, tenderebbe a rappresentare solo "l'uso linguistico proprio della rete", ovvero un uso specifico, piuttosto che una variazione. Inoltre, l'informazione linguistica risulta a suo modo "sporca" ed eterogenea, in quanto nelle pagine html convivono informazioni non necessariamente appartenenti allo stesso testo. Per cui la stessa idea di "ipertesto" non è adeguata a identificare le unità minime dei campioni di un corpus, ovvero i "testi". Da questo punto di vista la sola informazione linguistica tratta dalla rete non sarebbe adatta a rappresentare l'universo della comunicazione linguistica e rappresenterebbe solo se stessa.

Questi problemi, unitamente a problemi di carattere specificamente computazionale, debbono essere affrontati, e sono in effetti stati affrontati in vario modo nella costituzione dei web corpora nell'ultimo decennio, al fine di garantire criteri di rappresentatività.

Si deve notare però che l'espansione della rete che si è realizzata proprio in questi anni ha cambiato sostanzialmente i termini della questione come si erano posti sul crinale del millennio. La rete è attualmente l'ambiente privilegiato degli usi funzionali e culturali del linguaggio e contiene testi di ogni genere (letterari, giornalistici, accademici, burocratici ecc). Questi testi sono sia pagine html sia testi originali in vari formati che sono comunemente diffusi attraverso questo canale. Inoltre, Internet costituisce, già ad oggi, il canale primario dell'uso linguistico scritto, non solo a livello di fruizione, ma anche di espressione. In particolare la frequenza di utilizzo da parte della cittadinanza della lingua scritta è aumentata enormemente in connessione alla disponibilità della comunicazione in rete.

L'enorme varietà delle fonti di rete può quindi essere sfruttata per rappresentare l'uso italiano e la centralità di questo canale per l'uso linguistico nell'universo di riferimento garantisce della sua rappresentatività. Non corrisponde più ai fatti l'idea che la rete rappresenti se stessa, anche se non è possibile pensare al web come ad un corpus in se stesso senza rinunciare ad un criterio di campionamento.

3.2.2 Le risorse italiane accessibili in rete che è possibile utilizzare per ricerche linguistiche *online* sull'uso italiano sono già molto significative (Barbera, Corino & Onesti 2007). Segnaliamo qui, per la loro particolare rilevanza, cinque risorse: CORIS/CODIS, Corpus la Repubblica, NUNC, WEBBIT, ItWaC. Tali risorse costituiscono una grande ricchezza per la conoscenza e la diffusione della lingua italiana e la pongono tra le realtà linguistiche meglio rappresentate per quanto riguarda i corpora in rete.

Il corpus CORIS (CORpus dell'Italiano Scritto contemporaneo), realizzato presso l'Università di Bologna (Rossini Favretti et al. 2002), è una raccolta di testi scritti (principalmente degli anni '80 e '90) che si configura come un corpus di riferimento sul modello BNC. Il corpus generale costituito da circa 100 milioni di parole (a cui si aggiunge un corpus di monitoraggio) è composto da 38 milioni di parole di periodici, 25 milioni di parole di narrativa, 12 milioni di parole di prosa accademica, 10 milioni di parole di prosa giuridico amministrativa, da testi miscelanei per 10 milioni di parole e da una collezione di testi effimeri per 5 milioni di parole. Il corpus è accessibile parzialmente in modo libero e accessibile completamente per gli utenti accreditati. Da questo corpus è stata derivata una parte (CODIS, CORpus Dinamico dell'Italiano Scritto) rispetto alla quale possono essere selezionati sotto-corpora per ricerche focalizzate sui diversi tipi di lingua identificati nel corpus design o per operare ricerche e statistiche con diverse pesature delle componenti.

Il "Corpus La Repubblica", realizzato presso l'Università di Bologna Forlì, è un corpus di italiano giornalistico, liberamente accessibile dietro registrazione, che contiene di 350 milioni di parole tratte da articoli de La Repubblica tra il 1985 e il 2000. I testi sono categorizzati secondo genere ("notizie" e "commenti") e per una serie di generi tipici del contesto d'uso giornalistico ("religione", "cultura", "economia", "educazione", "notizie", "politica", "scienza", "società", "sport", "meteo"). La categorizzazione di genere è stata effettuata automaticamente attraverso tecnologie del linguaggio (analisi vettoriale, ).

I corpora precedenti sono accessibili in rete, ma non sono web corpora, in quanto collezionano testi digitalizzati che, alla sorgente, si avvalgono di altro canale di distribuzione.

Il NUNC "NewsgroupsUseNetCorpora" è un web corpus multilingue di 600.000.000 di parole per lingua, liberamente accessibile, specificamente dedicato alle pagine web relative a newsgroup tematici ("Cucina", "Motori", "Foto", "Cinema").

WEBBIT e ItWAC sono web corpora generalisti realizzati da Marco Baroni che sono rivolti alla rappresentazione della lingua attraverso metodi statistici di campionamento del web. WEBBIT è un corpus di 150.000.000 di parole liberamente accessibile, compilato con un campionamento di pagine ottenute attraverso *query* a un motore di ricerca. Il metodo prevede l'immissione nel corpus delle prime 10

pagine ottenute in risposta a una *query*. Le *query* sono state compilate attraverso una procedura standard (Sharoff 2006) di selezione e composizione di parole tratte da liste di frequenza. In particolare la procedura prevede la selezione di parole chiave con caratteristica semantica generale e non specifica (500 forme di parola frequenti) e 5.000-8.000 *query* contenenti 4 forme tra queste cinquecento casualmente generate e scaricate attraverso una infrastruttura di *crawling* dedicata alla formazione di corpora che sfrutta il *crawler* Heritix (BootCaT, Baroni & Bernardini 2004). Il WEBBIT è liberamente accessibile

ItWaC, probabilmente la risorsa principale a disposizione attualmente per l'italiano, è un web corpus realizzato in una primaria iniziativa europea dedicata alla formazione di web corpora (WaCky, *Web-as-Corpus kool ynitiative*, cfr. Baroni & Bernardini 2006; Baroni et al. 2009). Il corpus consta di 2 miliardi di parole tratte dal web in domini “.it” ed è generato attraverso il *crawling* di semi (*seeds*), ovvero siti web identificati attraverso Google in risposta a *query* formulate con 1000 coppie di parole di contenuto. Le coppie di sono derivate da una selezione di parole a media frequenza tratta dal “Corpus La Repubblica” e da una lista ulteriore appartenente al vocabolario italiano di base, da cui sono tolte le parole funzionali. I siti corrispondente ai semi sono stati interamente scaricati attraverso BootCaT. La strategia di formazione del web corpus si fonda sull'osservazione (Baroni & Ueyama 2006) che le *query* a Google con parole tratte da sorgenti quali i giornali ritornano materiali appartenenti alla sfera pubblica, mentre le parole appartenenti al vocabolario di base selezionano *blogs* e testi a carattere più personale. Il corpus ricavato può quindi aspirare a costituire un corpus di riferimento con un adeguata variazione. ItWaC è liberamente scaricabile ed è consultabile on line all'interno dell'infrastruttura Sketch Engine (Kilgarriff et al. 2004).

Complessivamente, i web corpora e i corpora in rete attualmente disponibili rappresentano ad un buon livello la generalità dell'uso dell'italiano, ma ancora in modo parziale la varietà di dominio d'uso della lingua.

3.2.3 Dal punto di vista del corpus design, RIDIRE va a colmare la mancanza di un grande archivio pubblico facilmente accessibile in rete che da un lato rappresenterà sufficientemente l'italiano nell'insieme delle sue varietà d'uso e dall'altro rende possibile selezionare il dominio d'uso di interesse dell'apprendente con un'ampia scelta tra quei domini d'eccellenza che caratterizzano la cultura italiana come primaria fonte di informazione culturale a livello globale. La base dati interrogabile attraverso il portale avrà una dimensione di circa due miliardi di parole e conterrà i dati relativi all'uso linguistico italiano in due tipi di domini:

A) domini semantici che identificano i campi dell'eccellenza italiana nel mondo per circa un miliardo di parole:

- Letteratura
- Moda
- Design-architettura
- Cucina
- Sport
- Religione
- Arti figurative
- Cinema
- Musica

B) domini nei quali la lingua si caratterizza per scelte legate al suo uso funzionale per circa un miliardo di parole:

- Informazione
- Economia e affari
- Amministrazione e legislazione

Il corpus è attualmente in corso di realizzazione. Sufficientemente grande e variato da poter costituire un corpus di riferimento per l'italiano<sup>8</sup>, diversamente da ItWaC non è ottenuto attraverso una selezione casuale di siti. La strategia adottata è di operare un *crawling* mirato su siti specificamente selezionati dalle unità di ricerca per la loro rilevanza e attinenza a ciascun dominio previsto. In altri termini, la finalità del *crawling* di RIDIRE è assicurare che le risorse scaricate dalla rete rappresentino con sufficiente certezza i domini in questione, a partire da siti selezionati sulla base di criteri espliciti. Ci si assicura cioè che le risorse contengano informazione linguistica caratterizzante ciascun dominio a livello della scelta lessicale e fraseologica. Passo essenziale questo per consentire lo sfruttamento dell'informazione linguistica da parte di apprendenti che intendono mettersi in grado di utilizzare le scelte canoniche della lingua italiana in specifici domini culturali o pratici. Su questa base, RIDIRE costituirà anche uno strumento di accesso alle sitografie delle risorse italiane di interesse culturale.

Proprio per facilitare la selezione dei domini di suo interesse da parte dell'utente e per massimizzare il requisito di attinenza dell'informazione linguistica al dominio RIDIRE è realizzato attraverso una infrastruttura di *crawling* e processamento dell'informazione scaricata che diminuisce il rumore nelle pagine scaricate dai siti oggetto di *crawling*.

---

<sup>8</sup> RIDIRE si configura come una infrastruttura dinamica e la dimensione di 2 miliardi di parole si riferisce alla prima *release*.

Anche il *crawling* di RIDIRE utilizza una interfaccia basata su Heritrix, che diversamente da BootCaT, si caratterizza per qualità tecniche necessarie al *crawling* mirato: (a) è una interfaccia di rete che consente il *crawling* distribuito e quindi l'inserimento in un unico *database* delle risorse scaricate in più sedi e in tempi diversi; (b) facilita anche ad utenti non esperti la selezione all'interno di un sito di quelle pagine che caratterizzano il dominio oggetto di *crawling* e l'esclusione delle pagine che forniscono informazione non caratterizzante il dominio o di carattere non testuale.

Le risorse scaricate subiscono un processo di elaborazione sequenziale (*pipeline*) nel quale il passaggio essenziale per rendere computabile l'informazione tratta dal web è la "pulizia del testo". Le pagine web presentano infatti quasi sempre testo non utile per scopi linguistici: pubblicità, link di navigazione, *credits* e in generale ciò che viene comunemente definito *boilerplate*. In RIDIRE questo testo viene rimosso attraverso due *tool* esterni (Readability e AlchemyAPI), che vengono eseguiti in cascata: se il primo non produce un risultato valido, viene attivato il secondo<sup>9</sup>.

Le risorse scaricate sono processate dunque da in una *pipeline* che comprende:

- pulizia dei file html e pdf e generazione di file di testo;
- tokenizzazione;
- annotazione di PoS e lemma con TreeTagger;
- filtro antiduplicazione;
- *language guessing*.

Le risorse scaricate sono quindi mappate in un *database* dal quale i testi scaricati possono essere validati. L'operatore può accedere al testo estratto, assegnare il metadato alle pagine che sono ritenute congrue o scartare la pagina.

Attraverso RIDIRE il web italiano sarà così considerato un corpus da cui estrarre la fraseologia utile a consolidare e apprendere l'italiano secondo metodiche moderne di estrazione dell'informazione da grandi moli di documenti.

### 3.3 RIDIRE per l'acquisizione dell'italiano L2

3.3.1 L'utilizzo dei corpora per l'estrazione di informazione linguistica si avvale di una serie di tecnologie che possono essere considerate standard. Il loro uso è

---

<sup>9</sup> La pulizia delle pagine web per la computazione del testo è un problema sensibile per le attuali tecnologie del linguaggio. Si veda Fairon et al 2007 e in particolare l'iniziativa di valutazione dei pulitori CLEANVAL. Per un dettaglio sulla strategia adottata in RIDIRE si veda Gregori et al. (in preparazione)

ampiamente presente in vari domini, primariamente in lessicografia, ma attualmente in modo esteso nelle tecnologie del linguaggio e nella cosiddetta *web technology*.

L'utilizzo dei corpora per l'acquisizione delle lingue seconde non è però immediato per gli studenti, e neppure per i docenti, e il loro uso in italiano L2 è una pratica ancora limitata<sup>10</sup>. La disponibilità di strumenti computazionali per l'estrazione di informazione viva da grandi corpora rappresentativi è però certamente una opportunità importante per l'acquisizione della capacità di utilizzo delle lingue seconde che è il portato naturale dell'evoluzione tecnologica<sup>11</sup>.

Ma cosa dovrebbe dare l'accesso al corpus ad un apprendente L2 per poter essere più significativo di un dizionario o di una grammatica? Ovviamente il corpus contiene l'informazione rilevante su una lingua e sul suo uso, che non è invece necessariamente presente né nel dizionario né nella grammatica e questa informazione è viva e non astratta. Quindi il corpus dà una certezza sulla realtà e l'appropriatezza dell'uso linguistico che istanzia, e non propone all'apprendente un processo deduttivo incerto, come avviene con gli strumenti tradizionali. In astratto, ovviamente solo in astratto, possiamo immaginare che nel corpus "la risposta c'è". Ad esempio, l'utilizzo del web come corpus da cui estrarre informazioni su lingue non conosciute attraverso *query* ai motori di ricerca è una prassi empirica che è divenuta comune tra i ricercatori. L'utilizzo dei corpora nel processo di apprendimento consente, in altri termini, di ipotizzare un *data-driven learning*.

Il corpus, però, è informativo se da esso si derivano dati sull'uso, ovvero se viene evidenziato ciò che in una lingua, o in un suo dominio specifico, si ripete, a livello delle scelte lessicali e sintattiche. Le concordanze sono lo strumento primario a questo fine (Sinclair 2003), ma pongono problemi importanti in ambiente di acquisizione L2, problemi che sono stati rilevati nelle varie esperienze succedutesi negli ultimi 15 anni<sup>12</sup>.

Le concordanze sono significative se danno immediatamente l'informazione sull'uso della parola o dell'associazione di parole che interessano, ma non sono funzionali se sono difficili da leggere e obbligano l'utente a cercare in pagine e pagine di contesti necessariamente frammentari. Esse pongono in particolare un problema di interpretazione dei contesti difficile da risolvere, specialmente se non si conosce bene la lingua oggetto.

Complessivamente le concordanze non sintetizzano l'informazione proposta. Ciò produce diffidenza negli apprendenti e negli insegnanti di lingua. Se l'utilizzo dei corpora obbliga all'acquisizione di metodiche troppo complesse e implica un

---

<sup>10</sup> Si vedano ad esempio Tognini-Bonelli (2000) Nicolás (2006) Hédiard (2009).

<sup>11</sup> Si vedano ad esempio Sinclair (2004); Conrad (2006), la principale conferenza di settore TALC e il recente seminario internazionale "New Trends in Corpus Linguistics for Language Teaching and Translation Studies", Granada 2008 (<http://www.ugr.es/~newtrends/>).

<sup>12</sup> La data di nascita di TALC è il 1994, il cui ultimo convegno internazionale si è tenuto nel 2010. Si veda in particolare Kilgarriff (2008) per una sintesi dei problemi in questo ambito.



dispendio di tempo eccessivo, si preferiscono gli strumenti tradizionali. Si dirà: “voglio apprendere una lingua, non la linguistica dei corpora”.

3.3.2 La conoscenza della coordinazione delle scelte lessicali secondo l'uso di una lingua (collocazioni)<sup>13</sup> è però necessaria al suo uso fluente e non è una informazione facilmente ottenibile senza il riferimento ad un corpus. Se considerata da questo punto di vista, la *Corpus Linguistics* non è un optional, ma una necessità. I documenti di rete collezionati in RIDIRE si propongono di costituire la sorgente della fraseologia italiana sia generale sia settoriale. In altri termini, attraverso l'infrastruttura sarà possibile dare all'utente evidenza dell'uso linguistico in un dominio e simultaneamente accesso ai suoi documenti, attraverso link alle risorse in rete. L'infrastruttura risponde quindi a entrambe le necessità sopra evidenziate: sarà insieme strumento di educazione linguistica e canale di accesso selettivo ai contenuti italiani della rete, quindi strumento di diffusione della cultura.

Ovviamente una massa di dati linguistici, che facilmente può raggiungere l'ordine dei miliardi di parole, è perspicua per l'apprendimento solo se le interrogazioni possono riferirsi a caratteri semantici e qualità linguistiche dei dati stessi, rispetto ai quali i contenuti debbono risultare indicizzati e quindi interrogabili.

RIDIRE svilupperà una serie di algoritmi che forniranno, per la prima volta insieme in una risorsa italiana di rete, le seguenti informazioni sull'intero corpus o su suoi sottocorpora:

- concordanze;
- collocazioni;
- liste di frequenza;
- *keywords*;
- *colligations*;
- *PoS patterns*.

RIDIRE si conformerà alle migliori pratiche per quanto riguarda la computazione di dati linguistici in rete<sup>14</sup> migliorando lo stato dell'arte dei processi computazionali offerti nelle risorse linguistiche italiane. Ma come è stato evidenziato nel paragrafo precedente, perché le funzioni computazionali possibili siano sfruttabili effettivamente ai fini dell'acquisizione dell'italiano L2, l'infrastruttura deve presentarsi e fornire l'accesso ai risultati in un tempo e con modalità compatibili con le esigenze dell'utente. RIDIRE si pone questo problema in modo esplicito e sta sviluppando una strategia a suo modo peculiare, orientata allo sfruttamento dei dati derivati da corpora nell'ambito privilegiato della formazione universitaria, che

---

<sup>13</sup> Sinclair 1991.

<sup>14</sup> Si veda in particolare Sketch Engine.

consente una applicazione più realistica di strumenti con un certo grado di complessità. I requisiti secondo i quali l'interfaccia di rete sarà sviluppata saranno in particolare:

- semplicità di utilizzo delle funzioni di ricerca;
- delimitazione dell'ambito di utilizzo privilegiato a livello universitario;
- disseminazione della conoscenza dell'infrastruttura e delle sue potenzialità nella comunità degli insegnanti di italiano all'estero e degli insegnanti italiano L2 in Italia;
- orientamento della principale funzione dell'infrastruttura al potenziamento dell'uso italiano nella sua varietà scritta;
- orientamento alla adeguatezza di utilizzo dell'italiano in domini semantici e funzionali determinati;
- riduzione, ma non azzeramento, dei pre-requisiti concettuali per l'uso, possibile appunto a livello universitario;
- determinazione chiara delle esigenze tipiche degli apprendenti a cui le ricerche possono rispondere e dei limiti delle risposte ottenibili.

La base dati RIDIRE, integrata dagli strumenti di computazione dell'informazione linguistica, inseriti in un contesto infrastrutturale dedicato specificamente a risolvere problemi acquisizionali, permetterà quindi un accesso selettivo alla fraseologia italiana e costituirà la sorgente d'informazione necessaria per consolidare le capacità di effettivo utilizzo dell'italiano da parte degli apprendenti. Questi, e i professori di lingua e cultura italiana nel mondo, e in generale i soggetti che vogliono potenziare le loro capacità nell'uso linguistico italiano, potranno avere, attraverso servizi linguistici liberamente accessibili in rete, informazioni sulla fraseologia specifica dei domini d'uso linguistico propri della cultura e della vita sociale italiana che sono rappresentati nel web corpus.

## Bibliografia

- Atkins, S., Clear, J. & Ostler, N. 1992. Corpus Design Criteria. *Literary and Linguistic Computing*, 7(1): 1-16.
- Barbera, M., Corino, E. & Onesti, C. 2008. *Corpora e linguistica in rete*. Perugia: Guerra Edizioni.
- Baroni, M. & Bernardini, M.S., 2004. BootCaT: Bootstrapping corpora and terms from the web. *Proceedings of LREC 2004*.
- Baroni, M. & Bernardini, M. S (eds) 2006. *Wacky! Working papers on the Web as Corpus*. Bologna: GEDIT. <http://wackybook.sslmit.unibo.it/>

- Baroni, M., Bernardini, M.S., Ferraresi, A. & Zanchetta, E. 2009. The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. *Language Resources and Evaluation* 43(3): 209-226.
- Baroni, M. & Ueyama, M. 2006. Building general- and special-purpose corpora by Web crawling. In *Proc. 13th NIJL International Symposium, Language Corpora: Their Compilation and Application*, Tokyo, 31-40.
- Blanche-Benveniste, C. (coo.) 1997. *EuRom4. Méthode d'enseignement simultané des langues romanes*. Firenze: La Nuova Italia Editrice.
- Burnard, L. (ed.) 2000. *Reference Guide for the British National Corpus* (World Edition) <http://www.natcorp.ox.ac.uk/archive/worldURG/urg.pdf>
- Conrad, S. 2006. Challenges for English Corpus Linguistics in Second Language Acquisition Research. In Y. Kawaguchi, S. Zaima, T. Tackagaki, Y. Tsuruga, M. Usami (eds) *Linguistics Informatics and Spoken Language Corpora*. Amsterdam: Benjamins.
- Fairon, C., Naets, H., Kilgarriff, A. & de Schryver, G.M. (eds) 2007. *Building and Exploring Web Corpora: Proceedings of the third WebasCorpus workshop, incorporating CleanEval*. Lovanio: Presses universitaires de Louvain.
- Gerrand, P. 2007. Estimating linguistic diversity on the Internet: A taxonomy to avoid pitfalls and paradoxes. *Journal of Computer-Mediated Communication*, 12(4), article 8. <http://jcmc.indiana.edu/vol12/issue4/gerrand.html>
- Gregori, L., Panunzi, A., Fabbri, M., Moneglia, M. & Paladini, A. (in preparazione). Cleaning HTTP and PDF files for the RIDIRE web corpus. Evaluation of results.
- Hédiard, M. (ed.) 2009. *Linguistica dei corpora: strumenti e applicazioni*. Cassino: Edizioni Università di Casinò.
- Ide, N., Reppen, R. & Suderman, K. 2002. The American National Corpus: More than the Web can provide. In *Proceedings of the Third Language Resources and Evaluation Conference (LREC 2002)*, 839-844.
- Kilgarriff, A. 2001. The Web as corpus. In P. Rayson, A. Wilson, T. McEnery, A. Hardie & Sh. Khoja (eds), *Proceedings of Corpus Linguistics 2001*. Lancaster (UK) : University Centre For Computer Corpus Research on Language, 342-344.
- Kilgarriff, A. 2008. Corpora in the classroom without scaring the students. Paper presented at the Conference New Trends in Corpus Linguistics for Language Teaching and Translation Studies. In Honour of John Sinclair, Granada 22-24 settembre 2008.
- Kilgarriff, A., Rychly, P., Smrz, P. & Tugwell, D. 2004. The Sketch Engine. In *Proceeding of the 11th Euralex Conference*, Lorient (France) 6-10 July 2004, 105-116.
- Nicolás, C. 2006. *Ricerche sul Corpus del parlato romanzo C-ORAL-ROM. Studi linguistici e applicazioni didattiche per l'insegnamento di L2*. Firenze: Firenze University Press.
- O'Neill, B.F., Lavoie, R. & Bennett 2003. Trends in the Evolution of the Public Web 1998-2002. *D-Lib Magazine* 9(4). <http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/april03/lavoie/04lavoie.html>
- Rossini Favretti R., Tamburini F. & De Santis C. 2002. CORIS/CODIS: A corpus of written Italian based on a defined and a dynamic model. In A. Wilson, P. Rayson & T. McEnery (eds) *A Rainbow of Corpora: Corpus Linguistics and the Languages of the World*. Munich: Lincom-Europa.

- Sharoff, S. 2006. Creating general-purpose corpora using automated search engine queries. In M. Baroni e M.S. Bernardini (eds), *Wacky! Working Papers on the Web as Corpus*. Bologna: Gedit, 63-98.
- Sinclair, J. M. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J. M. 2003. *Reading Concordances*. London/New York: Longman/Pearson.
- Sinclair, J. M. (ed.) 2004. *How to use Corpora in Language Teaching*. Amsterdam: Benjamins.
- Tognini Bonelli, E. 2000. Il corpus in classe: da una nuova concezione della lingua a una nuova concezione della didattica. In R. Rossini Favretti (ed.), *Linguistica e informatica: multimedialita', corpora e percorsi di apprendimento*. Roma: Bulzoni, 93-108.

## Sitografia

- AlchemyAPI. <http://www.alchemyapi.com/>
- BNC. <http://www.natcorp.ox.ac.uk/>
- BootCaT. <http://bootcat.sslmit.unibo.it/>
- CLEANEVAL. <http://cleaneval.sigwac.org.uk/>
- CORIS/CODIS. [http://corpora.dslo.unibo.it/coris\\_ita.html](http://corpora.dslo.unibo.it/coris_ita.html);
- Corpus La Repubblica. <http://dev.sslmit.unibo.it/corpora/corpora.php>
- Heritrix. <http://crawler.archive.org/>
- Internet World Stat. <http://www.internetworldstats.com/>
- ItWaC. <http://wacky.sslmit.unibo.it/doku.php?id=start>
- Linguasphere. <http://www.linguasphere.org/>
- NUNC. <http://www.corpora.unito.it/>
- Readability. <http://lab.arc90.com/experiments/readability/>
- RIDIRE Project. <http://lablita.dit.unifi.it/projects/RIDIRE>
- RIDIRE Crawler. <http://www.ridire.it>
- Sketch Engine. <http://www.sketchengine.co.uk/>
- TALC 2010. <http://talc2010.muni.cz/>
- TreeTagger. <http://www.ims.unistuttgart.de/projekte/corplex/TreeTagger/>
- WEBBIT. <http://clic.cimec.unitn.it/marco/webbit/>
- ZeitGeist. <http://www.google.com/intl/en/press/zeitgeist2008>

# ANCORA SULLE “RISORGENZE DIALETTALI”.

## PALERMO E IL SUO REPERTORIO FRA CONTINUITÀ E INNOVAZIONE

Mari D’Agostino  
Università di Palermo

### 1. Premessa

La Sicilia, la più vasta e popolata fra le isole del Mediterraneo, è una regione caratterizzata da una grande differenziazione interna dovuta sia a ragioni di ordine storico culturale che a vincoli ambientali. Alla tradizionale contrapposizione fra Sicilia interna e Sicilia costiera, si preferisce oggi una visione più articolata e composita che comprende almeno una tripartizione in «perimetro urbano costiero», «regioni-cerniera», «nodi urbani e contesti insediativi dell’interno» (De Spuches, Guarrasi & Picone 2002).

Il perimetro urbano costiero (che comprende l’area della costiera ionica dalla zona di Taormina a Siracusa, la cuspide della costa meridionale da Sciacca a Modica, il sistema metropolitano di Palermo e la fascia costiera da Marsala a Custonaci) è il più dinamico sotto il profilo demografico ed economico. Persistenti segni di criticità, che si manifestano anche con il declino demografico e l’abbandono da parte della popolazione più giovane, presentano generalmente i contesti urbani dell’interno, pur registrandosi una significativa differenza fra centri capaci di esercitare una funzione di controllo e di promozione del territorio circostante (Caltagirone, Canicatti, Enna, Caltanissetta) e aree caratterizzate da marcato isolamento e spopolamento.

Funzione di raccordo fra l’una e l’altra area è svolta dalle cosiddette regioni cerniera come l’area di Sciacca, Castelvetro, Alcamo che, dotate di una particolare dinamicità, propagano verso l’interno «stili di vita, relazioni sociali e innovazioni economiche tipiche della realtà urbane del perimetro costiero» (De Spuches, Guarrasi & Picone 2002: 134).

## 2. Assetto Sociolinguistico. Le indagini quantitative dal 1985 ad oggi

Il repertorio linguistico della Sicilia comprende, oltre all'italiano regionale, le diverse varietà diatopiche del dialetto siciliano, le parlate galloitaliche, le parlate siculo-albanesi. L'Isola è una delle regioni italiane meglio esplorate dal punto di vista sociolinguistico grazie anche ad indagini quantitative che si sono susseguite dal 1985 ad oggi. La prima di queste, denominata Osservatorio Linguistico Siciliano (OLS)<sup>1</sup>, «nettamente superiore, non fosse altro che per il grado di sottigliezza, delicacy per dirla alla Halliday, a ogni altra indagine consimile svolta in Italia» (Berruto 1992: 264), ci consente di avere una chiara idea, almeno fino agli anni '80, dell'organizzazione sociofunzionale del repertorio dell'Isola, delle sue dinamiche linguistiche e delle differenziazioni interne. Fra l'aprile del 1984 e il maggio del 1985 a circa 1320 residenti in Sicilia (estratti attraverso procedure casuali) è stato somministrato un complesso questionario sociolinguistico e di ognuno di essi è stato raccolto un brano di parlato in dialetto e uno in italiano attraverso la richiesta di descrivere due storie fotografiche. Dai dati raccolti emerge il quadro di una regione fortemente bilingue. Solo il 5,6% degli intervistati si dichiara, infatti, completamente dialettofono e il 3,6% totalmente italofono. Il 90% circa si ritiene, dunque, capace di percorrere liberamente lo spazio comunicativo a sua disposizione, riservando la selezione dei codici alle sole esigenze funzionali. Tali dati, che si riferiscono alla parte autovalutativa o metalinguistica dell'indagine, e che quindi devono essere interpretati con le cautele su cui più volte si è richiamata l'attenzione (D'Agostino 2007), vengono sostanzialmente confermati sia dalle indagini condotte a livello nazionale dall'Istat, come si dirà fra poco, sia dalle effettive realizzazioni linguistiche, testate sempre nell'intervista OLS attraverso la richiesta di costruzione di una storia in dialetto ed una storia in italiano, prendendo spunto, come si è detto, da due serie fotografiche. Benchè infatti, solo l'1% del campione si sia dichiarato incapace di costruire in dialetto la storia proposta (contro lo 0,1% in italiano), nel 12,5% delle interviste in dialetto (contro il 3,9% di quelle in italiano) il parlante ha avuto grandi difficoltà a mantenere il codice proposto esibendo una competenza assai limitata. Nell'interpretazione di questi dati non bisogna sottovalutare sia il ruolo della situazione intervista nel determinare tale sbilanciamento in direzione dell'italiano, sia le caratteristiche della prova (non si tratta di un parlato spontaneo), ma in ogni caso ben più dell'80% del campione ha dimostrato di sapersi muovere in forme diverse in entrambi i codici.

Il quadro d'insieme tracciato da questa grande ricerca quantitativa, può essere aggiornato servendosi delle indagini Istat 2000 e 2006 (Indagini multiscope) e dei

---

<sup>1</sup> I dati dell'OLS si riferiscono a un campione probabilistico di 1320 soggetti residenti in Sicilia e che avevano compiuto i 15 anni all'epoca della effettuazione dell'interviste (aprile 1984 - marzo 1985).

dati dell’Atlante linguistico della Sicilia (ALS), ancora in gran parte inediti. Tutti quanti confermano il volto di un’Isola nella quale, all’interno delle relazioni amicali e familiari, il dialetto conserva un significativo ruolo, mentre il monolinguisma italiano si riafferma come minoritario.

Quanti nel 2006 rispondono di parlare “solo o prevalentemente italiano” in famiglia sono infatti solo il 26,2% di contro al 72% circa che dichiara di parlare esclusivamente dialetto o di alternare i due codici. A circa 20 anni dall’indagine dell’OLS, e pur in presenza di notevoli difformità nella struttura del questionario che non rendono perfettamente comparabili i dati delle due inchieste (cfr. Lo Piparo 1990), lingua e dialetto sembrano essere saldamente presenti all’interno delle mura domestiche. I raffronti con l’analoga indagine dell’Istat, condotta nel 1988, sono da questo punto di vista assai significativi (vedi tabella 1). Gli scarti fra le due rilevazioni sono contenuti per ciò che riguarda i contesti “in famiglia” e “con gli amici” dove la risposta “solo o prevalentemente italiano” viene scelta oggi da circa l’8% del campione in più rispetto a 18 anni fa. Una differenza veramente significativa è registrabile, invece, nelle dichiarazioni d’uso dell’italiano “con gli estranei”, laddove abbiamo nel 2006 un aumento di circa 19 punti percentuali rispetto al 1988, con il contemporaneo prosciugamento del serbatoio dei dialettofoni monolingui o quasi monolingui (come tali possono essere interpretati a grandi linee coloro i quali dichiarano di parlare anche con gli estranei “solo o prevalentemente dialetto”).

Tabella 1. Sicilia linguistica. Confronto fra i dati Istat del 1988, del 2000 e del 2006

<b>Tipo di linguaggio usato</b>		<b>1988</b>	<b>2000</b>	<b>2006</b>
<b>In famiglia</b>	Solo o prevalentemente italiano	18,2	23,8	26,2
	Solo o prevalentemente dialetto	48,0	32,8	25,5
	Sia italiano che dialetto	33,0	42,5	46,2
<b>Con amici</b>	Solo o prevalentemente italiano	21,0	28,4	30,5
	Solo o prevalentemente dialetto	37,9	26,6	19,1
	Sia italiano che dialetto	39,9	44,2	48,7
<b>Con estranei</b>	Solo o prevalentemente italiano	40,7	57,1	59,1
	Solo o prevalentemente dialetto	25,7	12,7	9,8
	Sia italiano che dialetto	32,3	29,4	29,7

Questi dati ricevono sostanziali conferme dalle grande mole di risultati dell’Atlante linguistico della Sicilia, attualmente in corso di sistematizzazione. L’ALS è un atlante di repertorio che intreccia fortemente strumenti e metodi delle linguistiche della variazione ed è caratterizzato da una grande attenzione a problemi di ordine metodologico (cfr. D’Agostino & Paternostro 2006) oltre che da una robusta componente informatica (cfr. D’Agostino & Ruffino 2005). La complessa rete di rilevamenti è costruita in modo da potere lavorare sia su Comuni singoli che su

reticoli di comuni, sia su famiglie (costituite da un nonna o un nonno, un genitore e una figlia o un figlio)<sup>2</sup>, che su parlanti presi nella loro individualità<sup>3</sup>. Anche dai recentissimi dati ALS si conferma come uno sguardo generale e strettamente quantitativo non può che rilevare la corposa presenza del dialetto nel repertorio dell'Isola.

### 3. Palermo: abbandono del dialetto e suo sdoganamento

All'interno di questo quadro è però importante segnalare la presenza di una serie di differenziazioni interne all'Isola che, già rilevate nell'indagine OLS, si sono accentuate negli anni. Una decisa tendenza verso il monolinguisimo italofono era già presente allora nelle fasce della popolazione giovane e istruita dei grandi centri urbani e in particolare di Palermo. I dati OLS mostravano infatti presso questa fetta di popolazione una forte spinta verso l'emarginazione della componente dialettale del repertorio, cosa che collocava i suoi comportamenti linguistici a grande distanza da quelli del resto della Sicilia. Questa dinamica di progressiva divaricazione viene confermata, e ulteriormente rafforzata, in una recente indagine relativa al comportamento degli studenti universitari di Palermo. Jolanda Scarpello (2009 e in stampa) ha sottoposto un articolato questionario sociolinguistico ad un gruppo di studenti dell'Ateneo palermitano confrontando le risposte dei nativi del Capoluogo di regione e dei fuorisede con quelle che a metà degli anni '80 erano state date dai laureati di Palermo e della provincia. Nelle tabelle seguenti (tratte da Scarpello 2009) vediamo le sorprendenti analogie fra i dati dell'OLS e quelli attuali.

---

<sup>2</sup> Ciascuna delle 5 famiglie è composta dall'incrocio di 3 livelli generazionali : NONNI (> 60) GENITORI (40-50) FIGLI (18-26) e da 3 livelli di istruzione (basso, medio, alto). Viene inoltre considerata la lingua di primo apprendimento (italiano vs dialetto). Inoltre per ogni centro si considerano due adolescenti individuati sulla base del livello di istruzione proprio e dei genitori.

<sup>3</sup> Gli obiettivi dell'ALS sono molteplici, da una parte esplorare le diversità linguistiche e metalinguistiche e le forme di variazione di natura geografica, sociale, generazionale, familiare, dall'altra fare emergere l'intreccio fra le dinamiche interne ed esterne al punto e la rete di rapporti che lega pratiche territoriali e pratiche linguistiche, l'organizzazione dello spazio e del territorio e la stratificazione linguistica. Il Corpus dell'ALS consiste attualmente in circa 4.000 ore di parlato registrato in interviste a circa 1300 intervistati omogeneamente distribuito nell'intera area regionale.



Tabella 2. Come parli con tua madre? Due inchieste a confronto

	Anni '80		Anni 2000	
	OLS Laureati Palermo %	OLS Laureati Provincia %	2007 Universitari Palermo %	2007 Universitari fuori sede %
<b>Italiano</b>	57	37	67	29
<b>Siciliano</b>	17	63	0	24
<b>Entrambi</b>	26	7	33	47

Tabella 3. Come parli con i parenti anziani? Due inchieste a confronto

	Anni '80		Anni 2000	
	OLS Laureati Palermo %	OLS Laureati Provincia %	2007 Universitari Palermo %	2007 Universitari fuori sede %
<b>Italiano</b>	54	22	55	18
<b>Siciliano</b>	20	73	8	52
<b>Entrambi</b>	26	5	37	30

Tabella 4. Come parli con gli amici? Due inchieste a confronto

	Anni '80		Anni 2000	
	OLS Laureati Palermo %	OLS Laureati Provincia %	2007 Universitari Palermo %	2007 Universitari fuori sede %
<b>Italiano</b>	57	31	43	26
<b>Siciliano</b>	3	32	0	11
<b>Entrambi</b>	40	36	55	63

Come si può vedere le dinamiche linguistiche che erano in atto a metà degli anni '80 a Palermo sembrano essere giunte, per così dire, a conclusione. Presso i giovani istruiti non vi è più alcun dominio che viene riservato, se non altro come spazio mentale, all'uso esclusivo del dialetto. Anche in riferimento all'interlocutore "parenti anziani", classicamente quello in cui più consistente è l'attivazione di una competenza dialettale, viene selezionata in maggioranza la opzione "solo italiano" mentre si è quasi azzerata la scelta "solo dialetto", ancora maggioritaria per i giovani universitari fuori sede. Come si vede lo scarto fra i due gruppi è analogo a quello che si era registrato 20 anni fa nelle inchieste OLS.

Molto interessante è inoltre il contesto "amici" dove abbiamo un dato in controtendenza. Si tratta infatti dell'unico ambito rispetto al quale, per entrambi i gruppi presi in esame, viene dichiarato un decremento dell'uso monolingue dell'italiano, rispetto a quello di 20 anni fa. La conversazione con gli amici si caratterizza infatti, a quanto ci dicono queste recenti inchieste quantitative, come l'ambito "per eccellenza" in cui italiano e dialetto convivono. La categoria "entrambi" è di per sé troppo vaga per dare una qualche indicazione sulle forme

assunte da questa compresenza che, come sappiamo, può avere caratteristiche assai diversi. Essa può significare banalmente “a volte parlo in italiano a volte in dialetto, a secondo di chi è l'amico” “passo dall'uno all'altro codice” “parlo in un codice con qualche termine nell'altro”, ecc. Prima di fare un passo in avanti abbandonando le indagini quantitative per guardare da una parte al dinamiche in atto in Italia, dall'altro a dati qualitativi e strettamente linguistici relativi alla Sicilia, è bene focalizzare la nostra attenzione su una ultima tabella che si riferisce alla prima lingua dei soggetti a cui stiamo facendo continuo riferimento. È questo l'ambito in cui i due gruppi di giovani di recente intervistati dichiarano una situazione molto simile: l'italiano occupa per entrambi lo spazio dell'oralità primaria.

Tabella 5. Come hai iniziato a parlare? Due inchieste a confronto

	Anni '80		Anni 2000	
	OLS Laureati Palermo %	OLS Laureati Provincia %	2007 Universitari Palermo %	2007 Universitari fuori sede %
<b>Italiano</b>	75	24,4	97	75
<b>Siciliano</b>	22	70	0	12
<b>Entrambi</b>	3	6	3	12

Se vogliamo quindi ricostruire a ritroso quello che fin qui i dati quantitativi ci indicano possiamo così schematizzare: il passaggio dal dialetto all'italiano come lingua materna, che negli anni '80 era avvenuto solo per le classi socioculturalmente più alte del Capoluogo di Regione si è esteso ai giovani istruiti dell'intera Regione, mentre ciò non è accaduto per il monolinguisimo italofono. Ciò che distingue i due gruppi è quindi il rapporto non con la lingua materna, ma con la lingua seconda. Essa sembra avere un ruolo significativo come codice autonomo solo nel gruppo dei fuorisede (vedi soprattutto il contesto “parenti anziani”). Tracce di ripresa di una qualche forma di dialettalità si trovano, anche per i giovani istruiti palermitani, nella conversazione con gli amici. Anche in assenza di altri elementi che prenderemo in considerazione fra poco, le autodichiarazioni relative agli interlocutori anziani (tabella.3) ci indicano che è assai improbabile che nel caso dei giovani universitari di Palermo l'uso alterno di lingua e dialetto (vedi l'opzione “entrambi”) dichiarato nella conversazione fra “amici” sia connesso a un cambiamento di argomento e di interlocutori. Per la grande maggioranza piuttosto in tale contesto viene percepita, e non censurata, una qualche componente dialettale (ad esempio *tag switching*, formule fisse, regionalismi, ecc.) nel proprio eloquio. Si tratta di giovani che si sono accostati al dialetto in fase di seconda socializzazione, al di fuori di famiglie in cui uno, o spesso entrambi i membri, avevano già l'italiano come lingua materna.

Molti dati qualitativi a nostra disposizione ci indicano che tali giovani palermitani presentano forme di competenza dialettale assai imperfetta tanto da intaccare anche il nucleo morfologico della lingua. Se vogliamo riprendere la nota

osservazione di Nancy Dorian per cui il “dialetto morirà con i suoi stivali morfologici”, che focalizza l’attenzione sul fatto che, anche in situazione di lingue morenti permane la conservazione di un nucleo morfologico ‘autonomo’, presso questi giovani, che pure dichiarano di usare il dialetto, questo codice è già, se non morto, moribondo. La figura 1 rappresenta, ad esempio, uno degli striscioni esposti durante le grandi manifestazioni studentesche che si sono svolte a Palermo nell’Ottobre del 2008 ( conosciute con il nome di “onda anomala”, o più brevemente “onda” ). Lo slogan è semplice e ben costruito: “SUGNU SICILIANO VUOGGHIU STURIARI, L’UNIVERSITÀ UN L’AVITI A TUCCARI”. Dal punto di vista linguistico due sono gli elementi interessanti: il primo è la presenza del dittongo incondizionato -uo- tipico dell’area palermitana in “vuogghiu” (vs “vogghiu” di aree consistenti della Sicilia dell’interno), l’altro è la ‘abberrante’ presenza della desinenza -o- in ‘SICILIANO’ che denuncia la mancata competenza dell’estensore anche delle regole più elementari del dialetto che prevedono la terminazione in -u- dei nomi e aggettivi maschili.



Figura 1. Striscione esposto durante le amnifestazioni stuentesche dell’Ottobre 2008

Per comprendere un testo come quello esaminato dobbiamo inserirlo in dinamiche nazionali che hanno visto in anni recenti l’allentamento della censura dialettale e l’emergere di “una nuova dialettalità” in ambiti quali gli SMS, la comunicazione

mediata dal computer (forum, chat, siti web, e-mail), la pubblicità, le lingue esposte (insegne dei locali, graffiti, ecc.)<sup>4</sup>, la musica giovanile. Ciò avviene sia in aree dove il dialetto mostra una grande vitalità, ma anche laddove la sua perdita di parlanti, funzioni e strutture linguistiche è stata più radicale. Berruto (2002: 48), a questo proposito, rilevava che «un motto dell'Italia alle soglie del terzo Millennio sembra essere 'ora che sappiamo parlare italiano, possiamo anche (ri)parlare dialetto' ». Si tratta dunque, secondo Berruto, di fenomeni che vanno interpretati, anzitutto, come segnali di una definitiva italianizzazione della nazione, tale da consentire anche una nuova collocazione del dialetto, non più visto come sintomo d'ignoranza e ostacolo alla promozione sociale.

Uno striscione come quello degli universitari dell'onda, dunque, fa parte di, e a sua volta crea, un clima complessivo di allentamento della censura sociale nei confronti delle varietà locali (si è parlato a questo proposito di un loro 'sdoganamento'). Di questa nuova dialettalità partecipano, come indicano molte ricerche, anzitutto i giovani. Come avviene in altre aree della nazione a Palermo, rilevando la forte presenza del dialetto in SMS, scritture esposte, chat,<sup>5</sup> non possiamo che intravedere dietro ad essi soggetti che in parte comprendono ma che non parlano dialetto, individui che trovano in queste nuove forme di comunicazione un luogo dove usare una lingua che non hanno mai veramente acquisito senza preoccuparsi della norma e dell'errore. Sono forme di attivazione incompleta, frammentaria, e per alcuni aspetti caotica, che difficilmente potranno evolversi in

---

<sup>4</sup> Da studi condotti anche in altre realtà nazionali (Inghilterra, Francia, Germania) emerge che tali modalità comunicative favoriscono il contatto e la mescolanza linguistica. Si tratta anzitutto, come si è detto, di contaminazione fra caratteristiche proprie dell'oralità e della scrittura (si pensi al tentativo di riprodurre gli aspetti non verbali della comunicazione attraverso gli *emoticons*, cioè le faccine), ma anche della corposa presenza di prestiti, ibridismi, alternanze di codice. Tali lingue 'meticce' traggono la loro originalità sia dal substrato tecnologico, sia dalla situazione della comunicazione, oltre che, in maniera essenziale, dalla biografia sociolinguistica dei soggetti coinvolti. Ricerche recenti hanno esplorato alcune realtà italiane confermando le dinamiche linguistiche individuate in altri paesi. Negli SMS che si scambiano i ragazzi italiani, ad esempio, è ampiamente attestato l'utilizzo di varietà miste italiano/dialetti a cui, non di raro, si aggiungono lingue straniere (inglese, spagnolo, francese soprattutto).

<sup>5</sup> Si è variamente spiegato perché esso si trovi proprio in questo tipo di produzioni linguistiche. Moretti 2006 ritiene che siano rilevanti soprattutto tre aspetti: «la vicinanza alle modalità del parlato; la 'mancanza di tradizione' che lascia maggiore spazio a comportamenti innovativi da un punto di vista linguistico (presentando un minore controllo normativo); il carattere informale-scherzoso di molte comunicazioni, che fa sì che il dialetto diventi uno strumento importante di allargamento del 'potenziale di variazione' dei parlanti (cioè degli strumenti a disposizione dei parlanti per variare le loro modalità comunicative). A queste tre caratteristiche se ne può forse aggiungere una quarta, e cioè una 'volontà maggiore di riscoprire il dialetto', che sembra essere collegata alla minore presenza attuale di quest'ultimo rispetto al passato, alla notevole riduzione della sua conflittualità con l'italiano e non da ultimo ad un senso di 'nostalgia' rispetto alla tradizione, incrementato dalla sensazione che il dialetto non sia più vitale e solido nella società come un tempo» (Moretti 2006: 44).

una competenza più organica, riannodando, almeno in parte, i fili della trasmissione generazionale del dialetto che sembrano essersi spezzate nelle classi sociali alte di Palermo. L'utilizzo di un pseudo-dialetto consente solo, in buona sostanza, l'allargamento del potenziale di variazione, e non a caso viene autodichiarato in maniera assai rilevante nel contesto “amici” là dove la varietà di italiano comunemente usata, l'italiano giovanile, è particolarmente esposta all'innovazione e all'utilizzo di forme di mescolamento fra idiomi.

Assai diverso è il caso degli universitari fuori sede di Palermo (pure in gran parte con italiano come lingua materna) molti dei quali, come si è visto dai dati quantitativi, dichiarano di servirsi attivamente del solo dialetto nella comunicazione con anziani. Essi hanno in linea di principio aperte varie possibilità di discorso bilingue tutte quante, anche in questo caso, certamente favorite dall'allentarsi del pregiudizio dialettofobo. Scarpello (in stampa) segnala che in questa classe di parlanti è presente “una significativa varietà di gestione delle risorse bilingui. Se una scelta di posizionamento socio-linguistica più prossima ai coetanei istruiti di città determina un uso specialistico del dialetto e switching interfrasali o inserzionali sporadici a fini ludici, un'appartenenza comunitaria rivendicata, sostenuta dalla saldezza delle reti sociali di partenza, si traduce nella realizzazione di mixing fluenti bidirezionali, con ampi inserti interamente dialettali, da parte di quei gruppi di studenti che assumono un comportamento analogo agli adulti bilingui della micro-area di provenienza. In mezzo, a questi due gruppi, una varietà di usi e funzioni sociali del dialetto, come potenziale di variazione dell'italiano, con segmenti commutati più o meno estesi, al di sopra ma anche al di sotto della frase, accompagnati però quasi sempre da frequenti fenomeni di flagging che ne dichiarano l'intenzione pragmatica e la funzione epilinguistica”.

Particolarmente interessanti sono i casi in cui la dichiarata competenza bilingue è esibita non solo come un tratto chiave della propria identità ma anche come una marcia in più nel gioco dello scambio sociale e linguistico fra coetanei. Qui è Maurizio che parla, un ragazzo di 29 anni da poco laureato in Filosofia, e che proviene da un paese di una località di montagna, Gangi, a circa due ore di strada da Palermo. Le sue parole esprimono assai bene questo forte intreccio fra modalità comunicativa e identità linguistica:

Mm1: se ho | devo usare il mio linguaggio | se devo usare il mio dialetto e *cci hai infilari* | è na cosa ca fazzu in *continuationi puri ora* mentre stiamo parlando- io **passo** dall'italiano al dialetto / e anzi sta cosa qua mi piace proprio / quasi ostentarla con gli altri. però / chiaramente bisogna vedere in quali contesti ti muovi: non lo so. però è una cosa che ti viene in maniera naturale. di questa cosa ne vado anche orgoglioso perciò me la sento... (ex: tutti ridono)<sup>6</sup>

<sup>6</sup> Si riportano i simboli utilizzati: /, //, /// = pause più o meno lunghe; a: = allungamenti di diversa durata; abcd | efg = autocorrezione; abcd || efg = cambio di progetto; abcde+ = parola

Mm2: secondo me destavo anche curiosità nel momento in cui utilizzavo le mie frasi.

Rm1: *ncagliasti pròpriu i cristiani giusti tu!* (P: ride) [hai trovato proprio le persone giuste tu!]

Mm3: *unn'è ca ncagliàiu i cristiani ggiusti*, li ha conosciuti pure lei. (rivolgendosi alla sua ragazza) *un zu i cristiani titi, pipì*, assolutamente! [xxx] il punto è questo. partiamo dal fatto che io ho una certa autostima / io usavo le mie frasi in dialetto rendendole [ex: gli altri lo interrompono con un sovrapporsi di voci] e poi finisco *-picchi ia quannu pigliu a parola pui* (P: ride)- io usavo il mio dialetto in maniera così, naturale. ma poi quando c'era da parlare di qualunque tipo di cosa possibilmente / sapevo dire la mia in ogni caso. io ero quello che giocavo con il dialetto in maniera pure spudorata usando i termini più arcaici di questo mondo. però potevo parlare di musica, di politica, di qualsiasi cosa e *mi faccia ascutari puri ni atri maneri!* (ex: tutti ridono) [...] quando tu sei tranquillo è una cosa che trasmetti agli altri. se non sei insicuro gli altri non hanno cosa dirti [non è che ho trovato le persone giuste...non sono le persone ... perché quando prendo la parola poi!..e mi facevo ascoltare pure in altre maniere!]

Il comportamento linguistico di Maurizio e del suo coetaneo palermitano immaginario realizzatore dello striscione dell'onda anomala, sono comprensibili, come si è più volte segnalato, in un quadro di progressivo venir meno delle spinte dialettofobe in tutta Italia ma anche a Palermo, e in particolare nei suoi ceti socioculturalmente alti, dato questo per nulla scontato.

A metà degli anni '80 il Capoluogo di Regione si contrassegnava, rispetto a tutto il resto dell'Isola, per una chiara anomalia rispetto a quella regola generale della "serenità ideologico linguistica" per cui più alto è il livello di istruzione e più si proviene da un retroterra italofono, più si è disposti a viaggiare dall'italiano al siciliano e in particolare a non scoraggiare l'uso del siciliano nei figli, considerando questa seconda lingua ormai unicamente una fonte di arricchimento culturale (cfr. Lo Piparo 1990). Nei dati degli anni '80 il recupero in prospettiva, del dialetto siciliano, depurato totalmente dagli aspetti di "svantaggio sociale", vedeva, ovunque, in Sicilia i laureati in prima fila seguiti dai diplomati e via via dagli appartenenti ai livelli di istruzione più bassi. E ancora, fra i laureati, poneva al primo posto i più italofoni, dunque i residenti nei capoluoghi rispetto a quelli della provincia. Si tratta di una regola priva di eccezioni, tranne quella molto chiara e significativa dei laureati di Palermo. Essi la pensavano più o meno come i privi di titolo di studio della loro città, differenziandosi nettamente da quella che, semplificando, possiamo chiamare la classe dirigente del resto dell'Isola che invece spingeva molto avanti

---

interrotta; abcd. = intonazione conclusiva; abcd, = intonazione; [abcde] = enunciati sovrapposti; abc... = discorso lasciato cadere; (abcd) = commenti del trascrittore; \*abcd = parola ricostruita; [...] = espunzione; [x, xx] = elementi incomprensibili; abcd = elementi in siciliano. [abcde]= a fine turno i segmenti dialettali tradotti.

tale vocazione a un potenziale bilinguismo (vedi per questi dati D'Agostino 1996). Molti elementi indicano che a 20 anni di distanza sia stata Palermo a riallinearsi con il resto dell'Isola, e in particolare i giovani studenti universitari della città ad allentare i pregiudizi dialettofobi dei loro genitori. non significa ovviamente, come si è visto, che si possano riannodare i fili della trasmissione generazionale della lingua, ormai in gran parte interrotti. Si può forse inserire ibridismi e regionalismi in conversazioni con amici, usare il dialetto in funzione ludica e magari, guardare senza pregiudizi e forse con qualche ammirazione al coetaneo fuorisede che, invece, con le due lingue a disposizione sa giocarci veramente.

## Bibliografia

- Alfonzetti, G. 2001. Le funzioni del code switching italiano e dialetto nel discorso dei giovani. *Bollettino del Centro di Studi Filologici e Linguistici Siciliani* 19: 235-264.
- Auer, P. 2007. *Style and social identities. Alternative approaches to linguistic heterogeneity*. Berlin / New York: Mouton de Gruyter.
- Berruto, G. 1992. La Sicilia linguistica oggi. Un contributo importante alla sociolinguistica italiana. *Bollettino del Centro Studi Filologici e Linguistici Siciliani* 17: 255-59.
- Berruto, G. 2002. Parlare dialetto in Italia alle soglie del Duemila, in G. Beccaria, C. Mareello (eds), *La parola al testo*, Scritti per Bice Mortara Garavelli, Alessandria: Edizioni dell'Orso, 33-49.
- Berruto, G. 2005a. Che cosa ci insegna il 'parlare in due lingue'? Commutazione di codice e teoria linguistica e sociolinguistica. *Italian journal of linguistics* 17(1): 3-14.
- Berruto, G. 2005b. Convergence and mixing of language and dialects and models of language contact: the case of Italy. In P. Auer, F. Hinskens & P. Kerswill (eds), *Dialect Change Convergence and Divergence in European Languages*. Cambridge: Cambridge University Press. 81-95.
- Berruto, G. 2006. Quale dialetto per l'Italia del Duemila? Aspetti dell'italianizzazione e risorgenze dialettali in Piemonte (e altrove). In A. Sobrero. & A. Maglietta (eds), *Lingua e dialetto nell'Italia del duemila*. Galatina: Congedo editore, 101-127.
- D'Agostino, M. 1996. Spazio, città, lingue. Ragionando su Palermo. *Rivista italiana di dialettologia* XX: 35-87.
- D'Agostino, M. 2007. *Sociolinguistica dell'Italia contemporanea*. Bologna: Il Mulino.
- D'Agostino, M. & Paternostro G. (eds) 2006. *Costruendo i dati. Metodi di raccolta, revisione e organizzazione della banca dati nella sezione sociovariazionale dell'ALS*. Palermo: Centro di studi filologici e linguistici siciliani.
- D'Agostino, M. & Ruffino G. 2005. *I rilevamenti sociovariazionali. Linee progettuali*. Palermo: Centro Studi Filologici e Linguistici Siciliani.
- De Spuches G., Guarrasi V. & Picone M. 2002. *La città incompleta*. Palermo: Palumbo.
- Lo Piparo, F. (ed) 1990. *La Sicilia linguistica oggi*. Palermo: Centro Studi Filologici e Linguistici Siciliani.

- Moretti, B. 2006. Nuovi aspetti della relazione italiano-dialetto in Ticino. In A. Sobrero & A. Miglietta (eds), *Lingua e dialetto nell'Italia del duemila*. Galatina: Congedo editore, 31-48.
- Scapello, J. 2009. *Lingue, luoghi, forme della socialità. Gli universitari fuori sede di Palermo*. Tesi di Dottorato, Università di Torino, Palermo, Grenoble, Monaco, Neuchatel.
- Scapello, J. in stampa. La variazione del code switching nei discorsi di gruppo di giovani universitari siciliani. In R. Bombi, M. D'Agostino, S. Negro, R. Franceschini (eds), *Lingue e culture in Contatto*, Atti del X Congresso Internazionale dell'Associazione Italiana di Linguistica Applicata, AltLA, Bolzano 18-19 febbraio 2010).



# DYSPHEMISM AND NATIONAL, REGIONAL AND CLASS IDENTITY IN CONTEMPORARY ITALIAN

Nora Galli de' Paratesi

The American University of Rome

## **1. The evolution of euphemism and the emergence of dysphemism in Italian to-day**

Over the last few decades, the Italian language has gone through dramatic changes that have taken place extremely rapidly under the influence of modern media, causing what can be considered a linguistic revolution. Such changes have penetrated massively at all levels of the system and of sociolinguistic use. The aim of the present paper is to try to describe how such linguistic revolution has altered the sociolinguistic rules that should govern situation and register appropriateness and has in particular affected the balance between euphemism, taboo words and euphemistic substitutes. It also tries to analyse how such balance has been profoundly and perhaps irreversibly altered, bringing about major changes in the use of the Italian language.

In the past I carried out a study of euphemism (Galli de' Paratesi 1964), the result of which at the time showed a picture that is very different from the present one. Since then no other equally systematic research has been carried out on the subject and if we compare the language use described in that piece of research with the way speakers now relate with what were then taboo words vast differences emerge.

We can analyse the changes along three different dimensions: one is a comparison between the areas which were affected by verbal repression then as opposed to now, the second has to do with the evolution of the euphemistic substitutes (the words used instead of the taboo forbidden ones) and the third has to do with the strength of what has been defined as the “linguistic interdiction”, i.e. the psychological repression exerted on the speaker.

Many lexical items once prohibited are now in common use (for example: *casino* to mean “confusion”, *scopare* “to have intercourse”, *fregare*, *fregarsene* “not

to care”, etc.) and felt to be ordinary non taboo words, some other items that at the time were euphemisms are now dated and almost unusable because felt to be too “old fashioned”. Some examples are *in stato interessante* meaning “pregnant” or *passato a miglior vita* meaning “dead”. These two sets of examples point at two possible well known changes in the natural turnover of euphemisms. In the euphemistic lexicon there are taboo words that become less and less evocative of the prohibition they are struck by, less and less linked to the taboo object until they become “respectable” and can be used without embarrassment or offence. That is the case of *casino*, *scopare*, *fregare* and *fregarsene*, that are by now just low register, informal words used in a confidential situation, with a certain amount of humour attached to them but not taboo anymore. *In stato interessante* and *passato a miglior vita* on the other hand have become simply obsolete: in fact they were euphemisms and they were used to avoid more direct terms like *incinta* for instance. In the meantime *incinta* has been vastly used, it is not taboo anymore and it is the direct word normally chosen without embarrassment, and as a consequence *in stato interessante* is felt to be superseded and if anything signaling an embarrassment which is not necessarily there anymore because of a change in attitudes.

The weakening of the verbal interdiction is a general phenomenon in the evolution we are considering here, i.e. the changes occurred in the Italian language in the last three decades, but it varies according to the semantic field words belong to. The semantic areas around the concepts of death and illness are the ones in which a strong reservation still holds:  *cancro* is usually still substituted by *un brutto male* “a nasty illness” or *un male incurabile* “an incurable illness” and *morte* “death” is still a term to be avoided (the euphemistic metaphors *scomparsa* “disappearance”, *perdita* “loss” are commonly used). The sexual and scatological areas instead are the ones in which taboos used to be very strong and it is there that there has been the strongest change.

Any speaker of Italian that compares the contemporary use to my analysis carried out in the sixties would agree that such general weakening of the taboos is there and that it is striking and perceivable by anybody. In a recent study on political speech there emerges a language of politics that, compared to my original work quoted above shows a set of sociolinguistic rules concerning euphemism that are completely new (Bolasco, Giuliano & Galli de' Paratesi 2008). The ideal procedure in order to show the major changes that have taken places in the last decades in the linguistic use would be to devise an enquiry in order to measure the reactions of a sample of speakers to utterances that we know would have not been acceptable in the past. A thorough study of this kind is not possible here and therefore I have been forced to rely on my speaker's intuition. I have gathered a number of examples taken from newspapers and videos in which words that were taboo in the past appear to be freely used and taken for granted today by journalists, politicians and the general public. This intends to be a random sample of language in which the actual use of

taboo words would have been impossible in the past while nowadays no Italian speaker would find unfamiliar. The translation that is given tries to render the register level of utterances rather than the literal value of the single lexical items.

- Senator Marcello Dell'Utri: “che cazzo sta dicendo?” [what the fuck are you saying?] (to a journalist)<sup>1</sup>
- Journalist: “Ha detto così, l'abbiamo ascoltato tutti. Perché Mangano-mafioso dice il vero e Spatuzza-mafioso dice il falso?”  
Senator Marcello Dell'Utri: “Ma che cazzo dice! Da dove cazzo viene lei?” [What the fuck are you saying? Where the fuck have you been?]<sup>2</sup>
- Umberto Bossi, leader of the Northern League and Minister for Reforms: “Solo balle” [Just balls] (referred to the a witness in a Mafia trial)<sup>3</sup>
- Silvio Berlusconi, prime minister: “...un premier forte, duro e con le palle” [a strong and tough prime minister, with balls] (referring to himself during a discourse at PPEI)<sup>4</sup>
- Vittorio Sgarbi, art critic, former member of Parliament: “siamo un grande paese con un pezzo di merda come te” [we are a great country with a piece of shit like you] (during the talk show “Annozero”, addressee to a journalist)<sup>5</sup>
- Franco Zeffirelli, director: “Lei sta insultando una persona di valore. Questo è un teatro serio, vada via mascalzona, cretina. Lei è una stronza, vada a fare in culo: in culo!” [You are insulting an estimable person. This is a serious theatre, go away rascal, cretin. You are a piece of shit, go and get fucked: get fucked!] (to a journalist that was talking badly about the prime minister)<sup>6</sup>

---

<sup>1</sup> 4/12/09, <http://tv.repubblica.it/copertina/dell-utri-scontro-con-la-giornalista/39890?video>

<sup>2</sup> *Il Fatto Quotidiano*, 5/12/2009: 4.

<sup>3</sup> [http://www.ilgiornale.it/interni/il\\_pdl\\_il\\_fango\\_premier\\_danneggia\\_litalia\\_bossi\\_su\\_dellutri\\_spatuzza\\_solo\\_balle/spatuzza-gasparri-pdl-latorre-mafia-berlusconi-pentito-matteoli/05-12-2009/articolo-id=404505-page=0-comments=1](http://www.ilgiornale.it/interni/il_pdl_il_fango_premier_danneggia_litalia_bossi_su_dellutri_spatuzza_solo_balle/spatuzza-gasparri-pdl-latorre-mafia-berlusconi-pentito-matteoli/05-12-2009/articolo-id=404505-page=0-comments=1)

<sup>4</sup> 10/12/09, <http://tv.repubblica.it/dossier/processo-breve/berlusconi-c-e-il-partito-dei-giudici/40112?video>

<sup>5</sup> [http://video.corriere.it/?vxChannel=tuttiivideo&vxClipId=2524\\_7ee1140a-1830-11dd-a1a7-00](http://video.corriere.it/?vxChannel=tuttiivideo&vxClipId=2524_7ee1140a-1830-11dd-a1a7-00)

<sup>6</sup> *Il Fatto Quotidiano*, 11/12/2009: 14.

- Gianfranco Fini, President of the Chamber of Deputies in a meeting with youngsters between 8 and 18 declares that somebody who insults foreigners that come and work in Italy is “uno stronzo” [a piece of shit]<sup>7</sup>
- Umberto Bossi: “del PdL non me ne frega niente” [I don't give a fuck about the PdL] (referring to his coalition partner)<sup>8</sup>
- Vittorio Feltri: “Franceschini lo sfigato” [Franceschini the loser] (editorial title, referring to the leader of a party)<sup>9</sup>
- Silvio Berlusconi, Prime Minister: “Veltroni? Di Veltroni non me ne frega niente” [I don't give a fuck about Veltroni] (referring to the leader of a party)<sup>10</sup>
- A slogan during a centre-right demonstration: “Di Pietro, Bonino, Bersani, da sempre anti italiani e fuori dai coglioni” [Di Pietro, Bonino, Bersani forever against Italy off my arse]<sup>11</sup>
- Giorgia Meloni, Minister of Youth: “non me ne frega un cazzo” [I don't give a fuck] (answers to the security service that tries to hold her in order to give way to other people during a rally held by Berlusconi)
- Renato Brunetta, Minister of the actual government: “Abbandonate al destino questa élite di merda” [Leave this shitty élite to its destiny]<sup>12</sup>
- Berlusconi: “Se avete pazienza e culo avrei ancora un paio di cose da dirvi” [If you have patience and arse (‘luck’ in Italian) I have a couple of things to tell you] (in a rally)<sup>13</sup>
- Renato Brunetta, Minister of the actual government: “In Italia la sinistra è un'élite di merda, che vada a morire ammazzata” [In Italy the left is a shitty élite, it should go and get fucked]

---

<sup>7</sup> [http://www.ansa.it/web/notizie/rubriche/politica/2009/11/21/visualizza\\_new.html\\_1619484233.html](http://www.ansa.it/web/notizie/rubriche/politica/2009/11/21/visualizza_new.html_1619484233.html)

<sup>8</sup> [http://www.rainews24.rai.it/it/news\\_print.php? =76355](http://www.rainews24.rai.it/it/news_print.php? =76355)

<sup>9</sup> *Liberio*, 12/07/2009: 1.

<sup>10</sup> [http://quotidianonet.ilsole24ore.com/2008/10/07/123653-veltroni\\_siamo\\_pronti\\_collaborare.shtml](http://quotidianonet.ilsole24ore.com/2008/10/07/123653-veltroni_siamo_pronti_collaborare.shtml)

<sup>11</sup> [http://www.repubblica.it/politica/2010/03/20/dirette/roma\\_centrodestra\\_in\\_piazza\\_berlusconi\\_e\\_in\\_gioco\\_la\\_democrazia-2781891/](http://www.repubblica.it/politica/2010/03/20/dirette/roma_centrodestra_in_piazza_berlusconi_e_in_gioco_la_democrazia-2781891/)

<sup>12</sup> *Il Fatto Quotidiano*, 19/3/2010:13.

<sup>13</sup> *La Repubblica*, 28/09/2009.

- Renato Brunetta, Minister of the actual government: “Vada a morire ammazzata la sinistra che prepara il colpo di stato” [The left that is preparing a coup d'état should go and get fucked]<sup>14</sup>
- Declaration of a mayor of the Vicenza area: “Dobbiamo mettergli dei turaccioli in bocca e su per il culo a quei giornalisti” [We must put bungs in the mouth and up the arses of those journalists]<sup>15</sup>
- Daniela Di Sotto: “Io sono andata a sbattermi il culo con Storace” [I went to be bugged by Storace] (to her husband Gianfranco Fini, speaker of the Chamber of Deputies)<sup>16</sup>
- Phrases taken from telephone calls made by Berlusconi and by some managers of RAI, the National Broadcasting Corporation:  
 Innocenzi, undersecretary in the actual government: “Il capo mi fa il culo ogni 3 ore” [the boss is busting my arse every three hours]  
 Berlusconi to Innocenzi: “Che cazzo state a fare tutti quanti?” [What the fuck are you doing all of you?]  
 Innocenzi: “il problema vero è che poi ognuno va per i cazzi suoi” [the real problem is that everybody fucks around by themselves]<sup>17</sup>

The sources here are both written and spoken but the use of words that in theory should be taboo and that are used in an insulting way is equally abundant in both media. For the reader to appreciate how vast the use of these terms is not only in the political language but in general, I suggest that he or she Google for them: *non me ne frega niente* (“I don't give a fuck”) alone has 5.5 million hits appearing in various contexts. The fact itself that these phrases are used and are frequent in the political language confirms the speakers' intuition that they are part by now of the general language behaviour and all of us that live in this country, read the newspapers and watch television have no difficulty to recognize them as current.

However, as these examples show, the sheer size of their use in many different areas of linguistic interaction, in different media and levels of formality go beyond the physiological weakening of verbal repression which we have been speaking about before, they point towards a different phenomenon. What we have here is not the transition from euphemism to the straightforward use of transparent terms and the shedding of hypocrisy or reserve about an embarrassing content of the

---

<sup>14</sup> <http://www.repubblica.it/2009/09/sezioni/politica/berlusconi-divorzio-29/brunetta-sinistra/brunetta-sinistra.html>

<sup>15</sup> Stella G.A., *Negri, froci, giudei*, Milano, Rizzoli, 2009: 224.

<sup>16</sup> *Il Fatto Quotidiano*, 29/04/2010: 5.

<sup>17</sup> *Il Fatto Quotidiano*, 18/03/2010: 4.

conversation. What we have in fact is an obsessive use of terms many of which are in fact still strongly negatively connotated for many speakers, a phenomenon that linguists have started to study systematically from the '90s and which they called *dysphemism* (Allen & Buridge 1991) The normalisation of words like *casino* and the other ones we have examined, their becoming neutrally denotative and acceptable has nothing to do with the hyperuse of other words that still carry a strong prohibition and have a potential for offense (like *stronzo* "turd" in the sense of "bastard", *culo* "arse", *cazzo* "prick", *merda* "shit") and the use of which has very often just the function of insulting the addressee.

Dysphemism is an irrational use of language that is connected with insults, the semantic link between the lexical items and what they mean on one side and the context on the other is outside the usual linguistic conventional denotation. The semantic link is entirely illogical and connotative, words are there just to evoke heavy unpleasantness and cause offense.

In fact euphemism can have two opposites to it. One is when discourse is conducted without euphemisms, by the use of transparent terms without covering up the content of an utterance with clumsy and embarrassed allusions. In the natural evolution of languages many terms go from being embarrassing to being neutral, from prohibited and evocative to straight denotation. Even the most taboo topic has lexical items that are neither offensive nor necessarily negatively connotated and shameful. These items are the opposite of euphemisms, they are a straightforward way of pointing at an object. Examples of such direct use of words may be *cancro*, "cancer", which is in fact seldom used, *morire*, "to die" or *abortire*, "to have an abortion", instead of *un brutto male*, *mancare* or *avere un'interruzione di gravidanza*.

The other opposite of euphemism is dysphemism, the compulsive use (and hyperuse) of taboo words, the mirror image of repression, their obsessive, inescapable use. The compulsion is often so strong that words have no direct meaning but only an emotional value and they become insults, exclamations and terms of abuse as in the examples quoted above.

Here we find ourselves outside a rational, logical kind of communication and such behaviour is the result of a psychological phenomenon the root of which is the repression itself both of the concepts and of the words connected with them. The verbal prohibition transforms itself in a compulsion, which is as strong as the prohibition itself and which brings about an obsessive linguistic behaviour which is not connected by meaning to the taboo words.

It is precisely on dysphemism, its characteristics and its sociolinguistic reasons that we want to focus on in this analysis but in order to do that we must first describe the recent evolution of the sociolinguistic rules for the use of euphemism itself and the general transformation of the global phenomenon of verbal repression of which dysphemism is just one particular aspect.

## 2. Reasons for the weakening of verbal interdiction and the emergence of dysphemism

The reasons for the weakening of verbal taboo are to be found in the very deep changes through which Italian society has gone over the last decades. The general economic level in all social classes has grown as never before and as a consequence the whole class structure has changed. A vast urbanisation has taken place since the Sixties and there has been an unprecedented enlargement of education that has inflated the middle classes and reduced the numbers in the lower classes.

Larger opportunities for all have brought about more equality and diminished the social differences. The process of social change started with the so called economic miracle in the early Sixties and over two or three decades has altered the social structure of the country in many ways. On the linguistic level urbanisation, geographical mobility and higher education have favoured the spread of knowledge and use of the national language and weakened regional dialects. A more egalitarian social structure has wiped out the higher levels of formality in favour of a less formal, more direct approach between speakers in many situations that in the past might have been considered formal. A very significant example of this is again the political language. Until the beginning of the '90s the Italian ruling class was characterised by the use of a highly educated, even precious, certainly verbose and syntactically complicated language, as well as being encumbered by a kind of jargon that was almost esoteric and certainly full of terms that were hardly used in the common everyday language.

After the beginning of the nineties, the fall of the party system and the beginning of the so called "second republic" we have seen the birth of what is considered the new political language, that was initiated by two new political figures, by Umberto Bossi first and then by Silvio Berlusconi. The latter chose the theme of the "new" language as a sound bite and symbol of his new "movement" (Bolasco, Giuliano & Galli de' Paratesi 2008, Chapter 1): a new language for a new way in politics. The demise of the old political language, ironically termed *politichese*, and the appearance of the new one was immediately apparent both to the large public and to the newspapers and was studied by scholars of various disciplines, like political science, sociology and sociolinguistics. An equally large and profound change took place in the language used by both newspapers and television. A great contribution to this change was given by reality shows and other forms of popular television programmes that proliferated over the nineties.

Among the various characteristics of the new political language there are a more simple, paratactic syntax, shorter sentences and a less esoteric and learned lexicon, but the most important trait is the low level of formality that, starting from the language used on television, broadcasting and newspapers has conquered the use of

political speech in Parliament. Such informality manifests itself mainly as a real invasion by the spoken language into the written language and in the areas traditionally closer to the written medium that in the past was the level of reference of both political speech and the media, both oral and written.

The spread of informality in Italian is strictly linked with the demise of dialects. The use of the dialects by a large part of the speakers used to give many Italians a bilingual repertoire with various levels ranging from dialect to the national language through layers of subsystems of Italianized dialect and dialectal Italian. In fact the weakening of dialects has left empty the lower areas of the sociolinguistic register where dialects were traditionally used and Italian has moved into these areas. That means that younger speakers who are no longer proficient in the dialect now use Italian for their low formality level exchanges while before they would have used dialect and left Italian for the higher levels of formality. The diglossia situation in the past forced Italian to specialise as a formal variety and dialect as the informal language almost everywhere. Until the great waves of post-war urbanisation that started in the sixties the majority of Italians lived in small rural towns, in villages or directly on the land and they almost exclusively used the local dialects or strongly marked varieties. The national language was used for the written use except in Tuscany where the written language originally came from and partly in Umbria where the local dialect was typologically very close to Tuscan. Elsewhere it was only the educated middle class (at the time numerically not large at all) that could speak the language but would still most oscillate more or less largely between it and the local variety.

Now the number of Italians that have a monolingual repertoire is much higher. This fact means that for many speakers Italian (even if spoken in regional not yet fully standardized form) is by now strictly associated and vastly used for the first time with the most informal and expressive levels of communication for which before dialect was almost the only means. Therefore with the spread of monolingualism the low stylistic levels of Italian have assumed a stronger expressivity that was associated with dialect in the past. The informal registers of Italian that derive from this phenomenon have assumed the sociolinguistic rules of appropriateness typical of the dialects that have always been much less rigid than the ones of the language as far as euphemism, among other things, was concerned. And the speakers that are not yet monolingual and still typically oscillate between forms of Italianised dialect and dialectal Italian drag the sociolinguistic rules of dialect into the language.

The examples given above show a large use of words that were “prohibited” until recently. One way of looking at this phenomenon would be to analyse it in a diachronic dimension and show that for the older generation those words are taboo but they are not for the younger ones. It could be argued first sight that lexical items like *cazzo*, *fottere* (as in *non me ne fotte un cazzo*, “I don’t give a fuck” in practice



meaning “I don’t care at all”) have become current synonyms of *per niente*, *importare* (“not at all”, “to matter”) and or that *mettere/prendere in culo* are equivalent to “taking or losing advantage” and that these words have lost any taboo value or at least they have a very mild one and only for elderly people. But the picture is more complicated than that and the two polarity of embarrassment and defiance coexist and find themselves on a continuum that goes from the older speakers to the younger ones, along the class and education dimensions and vary also according to gender.

In fact what we have here is a massive linguistic change in progress concerning the sociolinguistic rules of appropriateness, a shift that involves the whole linguistic community and the whole language. The diachronic dimension of the language change is mirrored in the fact that the use and the psychological response to a large number of lexical items is the opposite in the two extremes of the age and social dimension and between the sexes. The use of expressions like the ones quoted above (words like *cazzo*, *stronzo*, *scopare*, *fottere*, *culo*, “cock”, “turd”, “to screw”, “to fuck”, “arse”) is normal for young males and banned for elderly educated women. At the moment they seem to communicate by the use of sociolinguistic subsystems that are very largely different, with the young people almost not trying to adapt to their interlocutors as it was mandatory to do in the past. That is to say that the prohibition still lingers with different degrees of intensity for all the speakers along these dimensions of variation. A proof of this is that when these words are used as insults people of all ages take offense and get angry which goes to show that they are still not acceptable and still possess a negative connotative potential. They are still charged for younger people as well, but younger speakers are motivated by new expressive urgencies. They want to use them *in spite and because* of their taboo potential and it is the prohibition itself that pushes them on the tip of their tongue. With time their hyperuse may most likely make their potential weaker, but there still is no doubt that they are very charged and if the old rules break down new ones shall have to be worked out to reestablish an acceptable social use.

### **3. Italian, dialect and national and regional identity**

In Italy the relationship between the dialects on one hand and the standard language on the other has always been a very troubled one, one might even say a very tense one, in the minds of the speakers since the unification of the country in 1860 and such tension has not yet allowed a complete acceptance of the national language.

As we have seen, because the national language has for centuries been much more of a written than an oral medium until the very recent but by no means yet complete disappearance of the dialects, it has remained until very recently associated

for many speakers to a high degree of formality and excluded from everyday interaction, at least in the majority of the regions.

In a statistical enquiry I carried out on the process of standardisation (Galli de' Paratesi 1984) I tried to focus on, among other things, the linguistic opinions of the speakers towards the standard language, the local varieties and the dialects. That study describes the state of affairs in the '70s when the dialects were still more vital than now (and the regional varieties still very close to dialect) and showed a strong rejection of the national language that was seen by the majority of the interviewees as a necessary tool for themselves and their children to try to reach success but an emotionally distant medium that belonged to others but not to themselves and that had been imposed on them at school and was therefore felt as very alien. The statistical data of the enquiry carried out then can still be applied to the contemporary situation. Here are some examples of the judgments expressed on the national variety by interviewees from Milan, Florence and Rome: *corretto* ("correct"), *esatto* ("exact"), *buono* ("good"), *pulito* ("clean"), *buon italiano* ("good Italian"), *istruito* ("educated"), *pronuncia neutra* ("neutral pronunciation"). Next to these positive judgments on the other hand we also have: *artificioso* ("artificial"), *non sciolto* ("unrelaxed"), *forzato* ("non spontaneous"), *ufficiale* ("official"), *astratto* ("abstract"), *non possono capirlo tutti* ("not everybody can understand"), *impersonale* ("impersonal"), *rileccato* ("dainty"), *affettato* ("affected"), *troppo perfetto* ("too perfect"), *non vivo* ("not lively"), *metallico* ("metallic"), *falso* ("false"), *innaturale* ("unnatural") (Galli de' Paratesi 1984: 191-196). The sample of the enquiry was selected by age, gender and socioeducational level and was representative of the population as a whole.

I can quote my own conclusions drawn then on the speakers' attitude as appropriate to the general feeling people show today when one asks them about their opinion on the standard language and the dialects:

"On the whole one can say that the Italians are torn between faithfulness to the national norm and loyalty to their own variety or dialect. One can talk of a polarisation of two opposite tensions: one, a sentimental one, towards one's regional speech and the other one, a rational one, a desire to own the means of sovraregional communication and of social promotion. They respectively are a centrifugal and a centripetal force in conflict with each other in the same individual. This seems to be the case more than in other unified nations. The stereotypes attached to the national standard seem to be largely negative ones of coldness, efficiency and lack of warmth... there is a real antipathy on the emotional level that clashes against a positive acknowledgment purely on the level of social prestige" (Galli de' Paratesi 1984: 202-3).

A natural expectation by speakers and scholars alike at the time when that enquiry was carried out was that once the national language became more largely used, this

state of affairs would change and the dialects would be shed in the standardisation process while the local varieties would take their place and become less and less marked and closer to standard Italian. The natural consequence was seen to be that the new generation of speakers would nourish different feelings towards the national language and make it their own as their linguistic behaviour was changing towards a local version of the national language.

As a matter of fact instead the linguistic behaviour has changed but the attitudes have not. The linguistic demise of dialects has in fact largely taken place as expected, with new generations of speakers that tend towards a passive knowledge of dialect rather than an active use of it and the establishment of the local varieties of Italian. More schooling and social ambition, as well as objective social conditions like social and geographical mobility and urbanization, have brought about the expected larger knowledge and use of Italian and a gradual loss of the dialects as well as the Italianization of the local varieties. However, the emotions are still very strongly partisan in favour of one's local origin and one's variety or even dialect.

The picture as a matter of fact amounts not just to the lack of national identity but to a hostile rejection of it. The reasons for this picture are deep in the history of the country.

The history of Italy has been a history of separation and in many cases of enmity between the small states in which the peninsula was divided. The most profound separation was between north and south. The first with a host of states some of which were independent for centuries and functioned for long periods as republican and democratic oligarchies. The south instead divided between the Papal States and the Kingdom of the Two Sicilies, both absolute monarchies. Both the social structure and the consequent mentality have always been different as was the economic development that made the two parts diverge more and more throughout history.

Soon after unification the rhetoric of the post-unification era was not enough to create the sense of a nation. The linguistic division for instance was so profound that the first world war was largely fought by citizens that spoke largely different dialects and varieties and could hardly understand each other and that were still immersed in their local culture. Nationalism later when it became associated with Fascism produced a strong and vacuous rhetoric that entirely disqualified. Its "imperial" pretensions and warlike aggressiveness were later felt as shameful, connected as they were after 1945 with a bloody dictatorship, with yet another terrible war and the defeat and ruin of the country. Fascism if anything alienated the idea of nationalism and as a consequence the concept and feeling of Italy as a unified country.

These divides have always made it difficult to accept the idea of one country and one language for all and the national language has been considered just the medium of the upper classes and of the privileged especially of Tuscan and of northern origin. Italy has always had a very strong populist component in its

political spectrum covering in practice more or less strongly all parties and stemming largely also from Roman Catholicism. Populism has recently been chosen as a defining trait by the two major groups of the right which together form the party in power at the moment. One of them in particular, the Lega Nord, has committed itself to racialism and a anti-immigrant and anti-southern policy, projecting the image of a strongly localistic, populist and very xenophobic movement that would like to reject any supposed external penetration. The fact that much of the "purity" of their northern stock is doubtful reveals its mythical nature when one realizes that a number of members of the Lega Nord are in fact second generation immigrants from the South.

The hostility towards a national language and a strong populist rhetoric add fuel to the powerful motor behind the use of dysphemism. It is not difficult to detect an attack to a very feeble and troubled national identity in the defense of the dialects as the "language of the people". It is the myth of the "true" local dialects with their irreverent and aggressive sociolinguistic rules as opposed to the "fake", "effeminate", "bloodless" and "antiseptic" standard language of the privileged classes. It is "us" against "them", the oppressed against the oppressors.

In the communities of the north of Italy where the Lega Nord is predominant in local government there have been a high number of initiatives in order to impose the use of dialect, from the local signs on the roads to their use in the assemblies of the local government and in the documents issued by the local offices. The most recent initiative is the inclusion of the use of the local dialect for shop signs together with other European languages with the exclusion of the non European ones (*Il Fatto Quotidiano*, 24/4/10: 2).

The use of a local dialect as opposed to the national language is also an attack against the immigrants. The national language acts as lingua franca and it is more often Italian (with local traits) that the immigrants learn, rather than the strict version of dialect that often nowadays the elderly only speak and the youngsters understand but in many areas do not use themselves anymore. By reintroducing the narrow form of the dialects (or at least its myth) the local population are affirming their most strict identity and cutting off the immigrants that have barely managed to master the national language.

Corrado Grassi (1964) studied this particular phenomenon in the sixties at the time of the big migratory movement from the south of Italy to the industrial centers of the north of the country. He noticed that at the time among the workers there was a revival in the use of the local dialect in Turin and he interpreted it as a means of marking the local identity against the newcomers who had to leave behind their own southern dialect and were resorting to their knowledge of Italian as a lingua franca to communicate with the local population. The dialect in this case becomes a social badge and a barrier to the immigrants' possibility to communicate and settle among the local population.

#### 4. Conclusion

We have seen that the reasons for the profound change in the use of euphemisms and the spreading of dysphemism are of various nature. They are the result of the profound social change brought about by decades of economic development. Wealth has brought about social and geographical mobility, urbanization and more education for everybody which are all factors of Italianization.

The first consequence of these changes is the weakening of dialects, the larger use of Italian in all social classes and the emergence and strengthening of spoken Italian in its various geographical varieties. In these varieties new sociolinguistic rules for the use of euphemisms have become established which are more suitable to a new social reality. In a new society with only marginal areas of diglossia Italian has become the language of almost all the citizens and a larger social equality allows more lenient rules of social respect while a new general level of informality makes words that once were taboo become acceptable.

But next to this tolerance for informality there seems to be also a general social intolerance for anything that represents something different from oneself, be it political opinions and ideas or provenance. Difference seems to engender a general social and political rage. Any political discourse, let alone discussion, by now seems to be inevitably tense and verging automatically onto insulting rather than arguing or even less reasoning, the level of conflict having gone far beyond the possibility of an exchange of ideas.

Such political rage has brought about a public language, the language of the media in all their pervasiveness, that has first percolated into and then flooded the language of general social exchange. This in its turn has accelerated the weakening of the sociolinguistic rules of formality that was in itself the positive consequence of a change towards a less unequal, more democratic society.

What we have now is not just a more modern, more egalitarian society, less encumbered by stiff formality, but a society which is by now profoundly ill. In it by now the informality is tantamount to lack of reciprocal respect and dignity of the other and lack of values in general. A society where politics without rules and respect of law is reflected in a lack of respect for the rules of sociolinguistic conventions that is the perfect mirror image of a political malaise.

## Bibliography

- Allen, K. & Burridge, K., 1991. *Euphemism and Dysphemism: Language used as shield and weapon*. Oxford: Oxford University Press.
- Bolasco, S., Giuliano, L. & Galli de' Paratesi, N. 2008. *Parole in libertà. Un'analisi statistica e linguistica dei discorsi di Silvio Berlusconi*. Roma: Manifesto Libri.
- Grassi, C. 1964. Comportamento linguistico e comportamento sociologico. A proposito di una recente pubblicazione. *Archivio Glottologico Italiano* 49: 40-66.
- Galli de' Paratesi, N. 1964. *La semantica dell'eufemismo. Eufemismo e interdizione linguistica nell'italiano contemporaneo*. Torino: Giappichelli.
- Galli de' Paratesi, N. 1984. *Lingua toscana in bocca romana. Tendenze verso l'italiano standard: un'inchiesta sociolinguistica*. Bologna: Il Mulino.

# LINGUA E IDENTITÀ.

## LA POLITICA NELLA RETE DI FACEBOOK

Marco Gargiulo  
University of Bergen

### 1. Introduzione

In questo lavoro sarà presentata la prima parte di un progetto di ricerca sulla lingua della comunicazione di argomento politico attraverso la rete. In particolare la ricerca riguarda la comunicazione attraverso i gruppi di discussione nei *social network* e tra i lettori dei quotidiani on line e pone l'attenzione non sulla lingua della politica in senso stretto, ma sulla comunicazione quotidiana di argomento politico, più precisamente la lingua degli interventi alle discussioni pubbliche in rete da parte di chi normalmente non si occupa di politica. Principale obiettivo di questa indagine sarà quindi quello di cercare di comprendere in quale modo il linguaggio dei politici ha imposto le regole di tale genere di comunicazione.

Come fa notare Gualdo (Gualdo & Dell'Anna 2004: 22) l'era di internet e dei nuovi media ha visto aumentare, nella mediatizzazione e nel marketing politico, la diversificazione e la varietà dei mezzi di propaganda; in questo campo lo strumento più potente è sicuramente rimasto la televisione, per la sua capacità di raggiungere il più grande numero di elettori nel più breve tempo possibile, ma ad essa si è affiancato, sempre più prepotente, l'uso della rete e delle varie ed efficacissime possibilità che essa offre alla pubblicità e alla diffusione delle idee. Il solo uso della posta elettronica permette, ad esempio, di raggiungere i destinatari nel cosiddetto tempo reale e di abbattere i costi della propaganda affidata alla posta tradizionale. Inoltre, attraverso canali di condivisione di video come *youtube* si può andare a (ri)vedere e (ri)ascoltare - ed eventualmente riproporre sulle proprie pagine internet - ciò che è stato trasmesso nei vari dibattiti politici disseminati tra le varie emittenti televisive pubbliche e private italiane; non solo, si può vedere ciò che in tv non è stato trasmesso, come registrazioni, amatoriali e non, di comizi, di dibattiti, di performance più o meno colorite. L'elettore più emancipato, inoltre, può navigare tra i siti dei partiti e andare a leggersi i programmi, le interviste, le dichiarazioni dei

candidati e i vari materiali contenuti negli archivi delle riviste e dei quotidiani di informazione on line.

Anche la politica italiana ha quindi invaso internet ed è sempre più forte e comune una vera interattività dei candidati con simpatizzanti ed elettori, dimostrata anche dalla crescita del numero di blog curati dai politici, poiché «all'avvicinarsi di ogni nuova scadenza elettorale i siti di partiti e associazioni si rifanno il trucco, adottando colori e immagini più accattivanti e corredano di registrazioni audio-video le loro pagine di rete» (Gualdo & Dell'Anna 2004: 23).

Silvio Berlusconi, il quale ha saputo utilizzare tutti i media in suo possesso per una vincente scalata e per il mantenimento del favore dell'elettorato, non ha però un blog attraverso il quale interagire personalmente e comunicare direttamente con i suoi sostenitori. Il mezzo con cui il leader del Popolo della libertà si trova maggiormente a proprio agio è, infatti, la televisione, confermando quindi quella immagine che lo vuole personaggio legato alla tradizione anche quando si tratti di comunicazione: resta emblematico dell'uso personalizzato dei media, il contratto che egli stipulò con gli italiani in televisione nel 2001.

I blog sono però creati, gestiti e molto frequentati dai suoi sostenitori, e, probabilmente, in maniera più massiccia, dai suoi detrattori: se si fa una ricerca tra i blog, inerendo il nome di Silvio Berlusconi attraverso il motore di ricerca Google, vengono fuori oltre 4890 risultati, mentre inserendo solo il cognome se ne ottengono 35.900. Tra i primi 20 risultati si trovano il sito ufficiale del Popolo della libertà; il sito istituzionale del Governo italiano ([governo.it](http://governo.it)); un blog contro ([noberlusconiday.it](http://noberlusconiday.it)) e uno a favore ([siberlusconiday.it](http://siberlusconiday.it)); segue una serie di blog che si pongono nei confronti dell'attuale Presidente del Consiglio in maniera critica o ironica e due siti antagonisti satirici, [silvioberlusconi.com](http://silvioberlusconi.com) e [berlusconisilvio.com](http://berlusconisilvio.com).

Anche Pierluigi Bersani, l'attuale segretario nazionale del Partito Democratico, e i suoi predecessori, Fabrizio Franceschini e Walter Veltroni, hanno il loro sito on line, mentre Gianfranco Fini si fa ospitare in quello del gruppo di Alleanza Nazionale ([alleanzanazionale.org](http://alleanzanazionale.org)) e Umberto Bossi in quello della Lega Nord ([leganord.org](http://leganord.org)).

## 2. La politica nella rete

Uno dei punti su cui convergono gli studi sul linguaggio della politica in Italia è il fatto che già da oltre un decennio, l'antipolitica, la mediatizzazione e la spettacolarizzazione del dibattito hanno contribuito, come causa e come effetto, alla semplificazione e alla banalizzazione del linguaggio. Giuseppe Antonelli (2000) parla, infatti, di passaggio dal paradigma della superiorità dei politici (la politica tradizionale italiana con la sua retorica umanistico-giuridica) al paradigma del rispecchiamento, attraverso il quale i nuovi politici hanno lavorato, qualcuno con



molta naturalezza e facilità, per offrire un'immagine di sé nella quale il cittadino potesse rispecchiarsi, proponendo modelli di comportamento il più possibile vicini a quelli della gente comune.

Dalla pagina del sito della Lega Nord si leggono le parole dei discorsi pubblici di Umberto Bossi, per certi versi molto lontane da ciò cui il cosiddetto *senatur* ci aveva abituati nella prima fase del suo apparire sulla scena politica, ma tuttora emblematiche di una sostanziale, ma evidentemente efficace, semplicità linguistica e di una retorica altalenante tra il sentimento identitario e il populismo<sup>1</sup>:

- (1) La Lega è come un bambino, è il frutto dell'amore. Io sono convinto che questo Movimento sia il risultato del lavoro generoso di migliaia di uomini e di donne che si vogliono bene, che vogliono bene alla città dove vivono, alla nazione cui sentono di appartenere. Il bambino è cresciuto, ha imparato a camminare con le sue gambe, ma bisognerà lavorare ancora perché diventi adulto e realizzi le sue ambizioni<sup>2</sup>.
- (2) La casa è dove c'è il cuore. Per noi è in Padania. Ci fu un tempo in cui non sapevamo neppure dove fosse finito il nostro cuore: in fabbrica, o in un'auto incolonnata in autostrada, o forse in una discarica industriale. Almeno adesso sappiamo che il nostro cuore è in Padania. Nell'antica Padania. Avevamo appena ritrovato la casa, quando gli schiavisti si sono affrettati a spiegarci che la Padania non è mai esistita e che noi padani eravamo condannati ad un esilio perpetuo e senza soluzione, peggio degli ebrei del "Và (*sic*) Pensiero". (Venezia, 21 settembre 2003).
- (3) Quando nel corso degli eventi umani diventa necessario per i Popoli sciogliere i vincoli che li legano ad altri, costituirsi in Nazione indipendente e sovrana ed assumere tra le Nazioni della Terra il ruolo assegnato loro dal Diritto Naturale di Autodeterminazione, il rispetto che si deve all'opinione della Società internazionale e dell'Umanità intera richiede che essi dichiarino le ragioni che li hanno costretti alla separazione. Da tempo memorabile abitiamo, dissodiamo, lavoriamo, proteggiamo ed amiamo queste terre, tramandateci dai nostri avi, attraversate e disetate dalle acque dei nostri grandi fiumi; Qui abbiamo inventato un modo originale di vivere, di sviluppare le arti e di lavorare; Noi apparteniamo ad un'area storica, la Padania, che sotto il profilo socio-economico è fortemente integrata al suo interno pur nella riconosciuta e rispettata diversità dei Popoli che la compongono [...] Noi Popoli della Padania [...] Noi siamo intimamente convinti che ogni ulteriore permanenza della Padania all'interno dei confini dello Stato italiano significherebbe lasciar spegnere lentamente ogni speranza di rinascita ed annientare l'identità dei Popoli che la compongono. (Venezia, 15 settembre 1996).

Come si vede anche nei pochi esempi riportati in 1-3, tra le forme della semplificazione usate dai politici, già analizzate dai linguisti e sociologi della

<sup>1</sup> Cfr. almeno Iacopini & Bianchi (1994) e Scarubbi (2005).

<sup>2</sup> Cfr.: <http://www.leganord.org/segretariofederale>

comunicazione<sup>3</sup>, sono raggruppabili diverse modalità comunicative: in primo luogo, aggressività verbale e infantilismo della comunicazione, messa in evidenza ad esempio da Nora Galli de' Paratesi (2004), analizzando i discorsi politici di Berlusconi dalla “discesa in campo” del 1994<sup>4</sup>; la studiosa ha inoltre messo in evidenza il passaggio da un linguaggio politico ampolloso, complesso e dotto della politica tradizionale a un linguaggio facile e immediato, «un linguaggio in cui il lessico è largamente quello di base, in cui i periodi sono molto più corti e molto meno complessi, con una sintassi paratattica molto meno gerarchica» (Galli de' Paratesi 2009: 139).

Proprio in tal senso, infatti, questa varietà di lingua semplificata va ad intaccare l'identità stessa della comunità che adopera quella lingua come punto di riferimento culturale e sociale; così non è difficile comprendere le parole dello scrittore Marco Mancassola (2006) quando, parlando della lingua della comunicazione politica italiana recente, afferma che:

il degrado linguistico di una nazione inizia quando le parole non appartengono più a nessuno. Quando entra in crisi l'idea che questa lingua è davvero la mia lingua, vive con me, respira in me, dipende da me, affonda nella mia carne e nei segreti del mio corpo. Quando ognuno parla come se quella lingua non lo riguardasse, non lo toccasse davvero, non lo sfiorasse neppure, esprimendosi per slogan anonimi e prefabbricati. Non personali, mai davvero importanti. È il vecchio incubo heideggeriano della chiacchiera o, più banalmente, il rovesciamento dell'utopia telelinguistica. Decenni dopo averci regalato un'appartenenza linguistica, la televisione italiana ha abolito tale senso di appartenenza. E chi la ama una lingua così, questa lingua-puttana che tutti scopano, ma nessuno sposa?

Un'altra caratteristica è, proprio per l'avanzata della lingua della quotidianità, è l'uso del dialetto che rientra sì nelle forme della semplicità, e come scrive Gualdo (2004: 27) funziona da «segnale di una vena più spontanea, più popolare e lontana dalla formalità ufficiale», ma ha anche lo scopo di creare un filo identitario e campanilistico con l'elettorato, il quale si riconosce non solo nel populismo e nella semplicità ma anche nel regionalismo di maniera.

Tutto questo ha invaso la rete, e i *social network* sono diventati le piazze virtuali in cui politici, commentatori appassionati ed elettori si incontrano, dialogano, fanno propaganda, si scontrano e dove *Facebook* è sicuramente il più popolare in questo momento. Nato come luogo in cui ritrovare amici di vecchia data, ex compagni di scuola e di università – dal nome del libro degli studenti dei college americani - ha poi inaugurato la sua funzione politica durante la campagna elettorale dell'attuale

---

<sup>3</sup> Si vedano ad esempio Leso (1994); Livolsi & Volli (1995); Calabrese (2002); Santulli (2005); Novelli (2006) e Gualdo (2006).

<sup>4</sup> Cfr. anche Bolasco et al. (2006).

Presidente degli Stati Uniti, il quale ha condotto un'abile raccolta di consensi anche attraverso la rete.

In Italia, sono stati pochi i politici ad aver usato finora *Facebook* come luogo privilegiato della propria campagna, ma molti hanno ormai una pagina personale o uno o più fan-club, per mezzo dei quali dialogano con i propri elettori, mantenendoli costantemente informati. Molti sono i gruppi in cui si discute di politica e spesso la creazione di tali gruppi è legata a fatti contingenti della realtà nazionale o locale, a questo o a quell'accadimento della movimentata scena politica e sociale italiana.

Il corpus raccolto per questo lavoro consta al momento dei commenti raccolti nelle pagine di una decina di gruppi di discussione che hanno per argomento la politica in senso ampio. I testi di cui esso si compone sono i testi brevi o brevissimi, che insieme compongono ipertesti multimediali e interattivi che sono, appunto, le pagine di discussione di *Facebook*.

Le pagine dei gruppi di discussione sono degli spazi aperti, creati da uno o più utenti e dedicati agli argomenti più disparati. Tra i più interessanti si trovano quelli di argomento politico, con l'invito a partecipare alla discussione rivolto a tutta la comunità degli internauti.

Come per ogni testo web, anche la lingua di questi post si caratterizza per la commistione di tratti tipici della comunicazione orale e scritta, oscillando continuamente da un versante all'altro, per competenze e capacità dello scrivente o per artificio retorico. Si vedano gli esempi riprodotti di seguito<sup>5</sup>:

- (4) Esimio Sig.X...evidentemente non ha letto attentamente i miei precedenti post...non mi duole questo...è cosa piuttosto comune qui dentro... Fortunatamente qui non tutti la pensano come Lei e ho la possibilità di confrontarmi e dialogare civilmente con persone che, a prescindere dal credo politico, apprezzano il confronto! Ossequi!»
- (5) ma chi ha voluto fare la simpatica? ho detto che siete come i fessi e gli ignoranti che parlano a vanvera e a sproposito..non c'è nulla di simpatico ma solo si constata la realtà..visto che non c'avete azzeccato nulla su di me!! sai, il sapientino lo si fa quando si hanno le carte in regole per farlo, altrimenti sono affari tuoi poi...mi sa che sei tu quello un pò agitato eh!..su avanti, calmo, prendi un pò di tavor e non dire tutte quelle parolacce che ad uno superiore come te non confà!!!!!!! vai tranquillo, continua a studiare l'italiano intanto.

Nell'esempio (4) è usata la strategia della lettera formale scritta per evidenziare la distanza che l'emittente vuole porre nei confronti del destinatario; nell'esempio (5), una forma dell'oralità controllata, in cui la punteggiatura ha lo scopo di riprodurre i ritmi e le pause del testo parlato. Non a caso i segni di interpunzione più usati sono i

---

<sup>5</sup> Per garantire l'anonimato ho provveduto a sostituire tutti i nomi con una X, rispettando comunque la grafia degli autori dei singoli interventi.

punti di sospensione – spesso a sostituire il punto fermo e la virgola - e i punti esclamativi e interrogativi, cioè i tratti più iconici a disposizione. Sono rari i due punti e il punto e virgola, mentre la virgola semplice è usata per lo più a segnare una pausa ritmica. La lingua, in generale, presenta i tratti tipici dell'italiano neostandard, di tipo informale e colloquiale, con frequenti cadute in direzione del substandard. Negli esempi seguenti si possono notare alcuni dei tratti tipici della semplificazione: punteggiatura limitata; uso della paratassi e di un periodare breve. Nella morfossintassi è tipico il sistema pronominale riorganizzato sul modello del parlato, come il ben noto impiego di 'lui', 'lei', 'loro' in funzione di pronomi soggetto in luogo rispettivamente di 'egli' / 'ella' / 'esso' / 'essa' / 'essi'; l'estensione dell'uso di 'te', pronomi tonico diretto, come pronomi soggetto; il predominare della forma pronominale indiretta 'gli' valida sia per il maschile e femminile singolare sia per il plurale; la semplificazione nell'uso dei pronomi relativi; ecc.

La testualità è caratterizzata dalla ricerca continua di espressività, attraverso un uso espressivo ed emotivo della punteggiatura, come abbiamo già visto, ma soprattutto, ed è la caratteristica principale di questi testi, di una sintassi estremamente semplificata, di alcune forme grafiche tipiche della scrittura giovanile e della scrittura trasmessa, come il maiuscoletto per rispondere alle esigenze di riprodurre le strategie espressive dell'oralità per surrogare i cosiddetti fenomeni soprasedimentali, come l'intonazione o il volume della voce; l'uso continuo di abbreviazioni attraverso troncamenti, contrazioni e sigle<sup>6</sup>.

Infine, tra le caratteristiche che qui mi preme evidenziare, l'uso di un lessico particolare e colorito, che spazia dal tecnicismo, all'elemento letterario ed arcaico, al neologismo, dai dialettalismi e regionalismi ai forestierismi più o meno adattati, dai gergalismi al turpiloquio, ai lemmi coprolalici e pornolalici, come negli esempi da (6) a (9).

- (6) Ok X accettiamo tutti i clandestini gli istruiamo, gli diamo tutto di quello che hanno bisogno pero poi devo tornare al loro paese a sistemare le cose in casa loro, in quanti tornerebbero nella propria patria??? Vuoi saperlo??? NESSUNO!!!! Xk una volta che si sono fatti una vita qui se ne fottono dei proprio connazionali, xk loro stanno bene e non hanno problemi...Intanto nel loro Paese proseguono le guerre e c'è sempre pronta a scappare...Entri in un circolo vizioso nel quale non risolvì niente...Allora meglio il motto della Lega: aiutarli si ma a casa loro...I volontari leghisti sono in Costa D Avorio dove hanno costruito un ospedale e pozzi...Ma queste notizie non ti arrivano, chissà xk??? Xk noi siamo quelli odiosi razzisti xenofobi...
- (7) non è razzismo ma è l'unico modo perchè non si scateni tra pochi anni un'altra guerra in italia.....NON VOGLIAMO CHE SUCCEDA IN ITALIA QUELLO CHE E' gia' successo in albania e in jugoslavia..... e poi diciamolo sinceramente i veri

---

<sup>6</sup> Cfr. anche Gheno 2009.



- (13) piu mi dicono che devo cambiare partito e piu rimango sempre convinto che il pdl e piu forte e libero che mai e piu infangano berlusconi dicendo che selafa con le veline e con noemi e piu mi convingo che votero sempre a lui ma pio anche con la nonna di noemi sefatto la foto allora anche a lei se fatta ahahaha che imbecilli sono sti comunisti ai forni e viva silvio meno male che silvo c'ee siiiii.
- (14) Bello kuesto gruppo, neonato e già ci sono i soliti pirla ke vogliono fare i perbenisti! ma mi domando e diko, se nn vi piace un gruppo KE KAZZO VI ISKRIVETE A FARE? SOLO PER ROMPERE IL BERLUSKONI ALLA GENTE PER BENE?? (kritika rivolto a kuella povera testina di Silvio ke ha postato alle 21.34) Per la kronaka kamerata Paolo X, io sono un fascio, ma allo psikonano lo prenderei a calci nelle gengive molto volentieri! Nn okkorre essere komunisti per odiare il nano, basta essere intelligenti e koerenti, forse tu nn lo sei e kuesto è un problema tuo, nn del resto del mondo, kuidni kiudi il koperkio ke hai sulla bokka e tira lo sciakkuone, la porta per uscire ed andare a fare in SILVIO la konosci bene, se nn la rikonosci è la stessa da kui sei entrato, kuindi alza i takki, vattene senza dare fastidio! SILVIO FACCIA DI KULO, NN PARLARE KE APPESTI L'AMBIENTE!

Altra caratteristica fondamentale è «l'abbassamento del livello di guardia dell'uso delle strategie eufemistiche» (Galli De' Paratesi 2009: 140) con l'abuso di espressioni provenienti dai registri più bassi della lingua e marcatamente violente ed aggressive, come negli esempi seguenti:

- (15) x fortuna in romania quando li arrestano succede ancora che li buttano nei sotterranei e li torturano giorno e notte quei zingari di merda, qui dovremmo fare lo stesso, zingari assassini ladri bastardi...
- (16) avete rotto i coglioni!!! perbenisti del cazzo!! voi e le vostre affermazioni del cazzo... solo perchè non avete i coglioni di guardare i problemi che gli extracomunitari creano!! gia vengono in Italia e fanno quel cazzo che vogliono, e per di più quando creano problemi li difendete!! siete voi che fate schifo, che sputate sulla terra che vi dà da Vivere! Sulla nostra Italia! se ora quà si vive da schifo è grazie a voi! e a che ha permesso questo scempio di Valori, Cultura e Patriotismo che anni e anni fa c'era! Vergonatevi!! W il DUCE!
- (17) Facciamo una cosa ???....Una bella nave, grande, riempiamola di comunisti (porelli), di extracomunitari, portiamola al largo e buttiamola a fondo....faremo solo del bene alla nostra cara Italia. I nostri nonni sono morti sul Carso, sul Piave, in Russia, in Africa , per dare un futuro a questa nazione, e loro che hanno fatto e fanno ?.....la stanno distruggendo, portando in Italia della immondizia umana ed ideologica (vedi i crocefissi che ci vogliono far togliere, delinquenza, assistenzialismo).....Gridiamo..Italia libera, o sara' guerra.....ma quella vera. !!!!

- (18) sta sicura che me fasso si e done straniere, che e xe anca più bee de ti che ti xe un sorse da magasen.

### 3. Conclusioni

Il lessico della politica degli ultimi quindici anni in Italia ha visto prevalere principalmente due campi semantici contrapposti e usati dai diversi schieramenti in condizioni differenti: nuovo vs vecchio; a questi poi si legano futuro vs passato, movimento vs staticità, positivo vs negativo.

Nelle discussioni politiche su *Facebook* si ha la contrapposizione, semplicistica e ovvia, tra (centro)sinistra e (centro)destra, tra berlusconiani e antiberlusconiani. Silvio Berlusconi è l'argomento prevalente della discussione mediatica, comparando in innumerevoli gruppi dai nomi più diversi: "SFIDA!", "STIAMO", "CERCANDO 20 MILIONI DI ITALIANI CHE NON VOTERANNO BERLUSCONI", "Scommetto di poter trovare 100.000 utenti che apprezzano Silvio Berlusconi", "Mobilitazione Nazionale per chiedere le Dimissioni di Berlusconi; Grazie Silvio", "Scommetto di poter trovare 1.000.000 di utenti che odiano Silvio Berlusconi", "Berlusconi nobel per la pace", "BERLUSCONI NON E' IL MIO PRESIDENTE", ecc.

L'attuale Presidente del Consiglio è apostrofato con vari appellativi, la maggior parte dei quali passa da un media all'altro: "psiconano", "nano malefico" e "nano di arcore", "Al tappone" - ormai frequentissimi, ma anche - "pornonano", "pornoduce", "porno puffo", "puffo", "nano piduista", "frottolo", ecc.

L'immagine che emerge da questa breve analisi è quella di un'Italia definita profondamente berlusconizzata e intrappolata nella rete dei media, un Paese più virtuale che reale, che trova conforto nel vedersi riprodotto nelle immagini televisive o sul web. Le discussioni infinite sui *social network*, o anche quelle che si trovano in calce agli articoli dei quotidiani on line - ogni testata giornalistica lascia infatti largo spazio ai commenti dei propri lettori - offre l'idea e l'illusione che la democrazia si eserciti nella rete del web e che si eserciti attraverso il linguaggio della violenza, attraverso l'imitazione di quello che Nora Galli De' Paratesi (2009: 143) ha giustamente definito come «linguaggio della sommossa da osteria, senza un progetto politico, che ha dato inizio tuttavia ad un modo nuovo di comunicare con gli elettori e di formulare lo scontento ed ha aperto la strada ad un abuso più sottile e pericoloso, quello in cui il bersaglio dell'insulto è costituito dai cardini della democrazia».

La lingua di questi post/spot fa pensare, più che ad un reale bisogno di comunicazione, ad un mediatico bisogno di partecipazione, di essere presenti nella piazza dove tutto accade; poco importa se la lingua perde significato, se col venir meno delle ideologie, anche le parole portano significati sempre più generici, se le

discussioni ruotano attorno alle accuse e alla vaghezza, se è immediato il passaggio all'offesa, anche violenta, o allo sberleffo sarcastico.

Persino parole un tempo direi sentite come sacre, come patria o libertà, in questa infinita bagarre collettiva, hanno un senso diverso, vuoto. Già Erasmo Leso nel 1994 proponeva una riflessione sulla storia della lingua politica partendo proprio dalle parole, e, citando il filosofo Aldo Giorgio Gargani, col fine di verificare l'autenticità dei nostri vocabolari e la genuinità dei nostri modelli di simbolizzazione, proponeva di riflettere proprio sul senso di parole fondamentali e simboliche come destra e sinistra, centralismo e federalismo, progressista e moderato. Quindici anni dopo l'era di mani pulite, se da una parte si lamenta l'allontanamento del cittadino dalla politica, dall'altra si ha il trionfo dell'antipolitica - altro termine abusato dai media e dagli attori del cosiddetto teatrino della politica e bandiera sotto la quale, a turno, pretendono di stare tutti - e, da un punto di vista semantico, il lessico della politica e della democrazia si svuotano di significato per rendere sempre più labile il legame della lingua con l'identità stessa di un popolo e della sua storia.

## Bibliografia

- Antonelli, G. 2000. Sull'italiano dei politici nella Seconda Repubblica. In S. Vanvolsem et al. (eds), *L'italiano oltre frontiera. Atti del V Convegno Internazionale*. Lovanio-Firenze: Leuven University Press – Cesati.
- Bolasco, S., Giuliano, L. & Galli De' Paratesi, N. 2006. *Parole in libertà. Un'analisi statistica e linguistica dei discorsi di Berlusconi*. Roma: Manifestolibri.
- Calabrese, O. 2002. *Come la boxe. Lo spettacolo della politica in Tv*. Roma-Bari: Laterza.
- Galli De' Paratesi, N. 2004. La lingua di Berlusconi. *Micromega* 1: 85-98.
- Galli De' Paratesi, N. 2009. *Eufemismo e disfemismo nel linguaggio politico e nell'italiano di oggi*. *Synergies Italie* n° special: 137-144. <http://ressources-cla.univ-fcomte.fr/gerflint/Italie-special/nora.pdf>
- Gheno, V. 2004. Prime osservazioni sulla grammatica dei gruppi di discussione telematici di lingua italiana. *Studi di Grammatica Italiana* 22: 267-308.
- Gheno, V. 2008. Il lessico dei newsgroup. Varietà di lingua a confronto. In Cresti E. (ed), *Prospettive nello studio del lessico italiano. Atti del IX Congresso Internazionale della Società di Linguistica e Filologia Italiana (SILFI)*. Firenze: Firenze University Press, I, 147-155.
- Gheno, V. 2009. I giovani e la comunicazione mediata dal computer: osservazioni linguistiche su nuove forme di alfabetizzazione. *Verbum Analecta Neolatina* XI/1: 167-187.
- Gualdo, R. 2006. Il linguaggio politico. In P. Trifone (ed). *Lingua e identità. Una storia sociale dell'italiano*. Roma: Carocci: 187-212.
- Gualdo, R. & Dell'Anna, M.V. (eds) 2004. *La faconda repubblica. La lingua della politica in Italia (1992-2004)*. Lecce: Manni.



- Iacopini, R. & Bianchi, S. 1994. *La Lega ce l'ha crudo! Il linguaggio del Carroccio dei suoi slogan, comizi, manifesti*. Milano: Mursia.
- Leso, E. 1994. Momenti di storia del linguaggio politico. In L. Serianni & P. Trifone (eds). *Storia della lingua italiana*. Einaudi: Torino.
- Livolsi, M. & Volli, U. 1995. *La comunicazione politica tra prima e seconda repubblica*. Bologna: Il Mulino.
- Mancassola, M. 2006. La lingua italiana dopo Silvio Berlusconi. <http://www.nazioneindiana.com/2006/03/23/lalinguaitalianadopsilvioberluscon>
- Novelli, E. 2006. *La turbopolitica. Sessant'anni di comunicazione politica e di scena pubblica in Italia: 1945-2005*. Milano: Bur-Rizzoli.
- Orletti, F. (ed.) 2004. *Scrittura e nuovi media. Dalle conversazioni in rete alla Web usability*. Roma: Carocci.
- Pistolesi, E. 2005. *Il parlar spedito. L'italiano di chat, e-mail e sms*. Padova: Esedra.
- Santulli, F. 2005. *Le parole del potere, il potere delle parole. Retorica e linguaggio politico*. Milano: Franco Angeli.
- Scarubbi, A. 2005. *La Lega qualunque. Dal Populismo di Giannini a quello di Bossi*. Roma: Armando.



# HUMOUR

## IN INTERCULTURAL PROFESSIONAL SETTINGS.

### A SHORTCUT TO LANGUAGE, COGNITION AND IDENTITY

Lita Lundquist  
Copenhagen Business School

#### 1. Humour – language, cognition, and national identity

This article will use humour as a focal point where language, cognition and identity meet. It will illustrate how a comparative analysis of the use of and reaction to humour in cross-linguistic and cross-cultural professional settings can help us shed light on the relationship between language, cognition and identity within a given language community, on the one hand, and help us explain differences between two language communities on the other. The languages studied in this case are Danish and French, and the ‘identities’ are the Danish and the French “national character” (von Humboldt 1822/2000) or “national culture”. Approaching these two notions with some caution, we shall regard a ‘national culture’ as “a collective programming of the mind”, i.e. «this component of our mental programming which we share with more of our compatriots as opposed to most other citizens.» (Hofstede & Usunier 2003: 137), and which lays down common ways of being, acting, thinking and referring to traditions and norms for a given society (d’Iribarne 2008: 9).

##### 1.1 Why humour?

Humour is a multifaceted phenomenon; this is exactly what makes it appropriate for the present comparative analysis of two languages and societies. It goes without saying that verbal humour includes language and language use; it builds on cognitive assumptions (see below) and provokes emotional effects (James 1884) in the form of pleasure and the physical reactions of smile and laughter (Chafe 2007), and it has huge social effects in its creation of new “momentary social ties” (Latour 2005: 64). This is why I shall consider humour to be a *mediator*, i.e. a movement or

a transformation which offers «occasions, circumstances [in which] lots of surprising aliens may pop up» (Latour 2005: 58). Such a mediator towards a new and possibly relaxed atmosphere is even more called for in cross-linguistic and cross-cultural professional encounters which are often governed by tense strategic aims and hampered by linguistic, cultural and professional barriers. Humour is also a social fact, necessarily involving two or more actors, and hence the use of humour is regulated by social norms prevalent in the given societies.

Below, I shall present important differences in how Danes and French use humour with each other, and try to explain these differences within both a linguistic-cognitive and a sociological framework, the last of which concentrates on how national identities have been shaped by specific civilisation processes (Elias 1994 /1939).

## 2. Humour used differently by Danes and French in professional settings

In a former study (Lundquist 2009), interviews with five Danes working in France and with five French people working in Denmark<sup>1</sup>, showed that there was consensus across both groups as to there being consistent similarities within the Danish and the French groups respectively, and differences between the two groups when it comes to using humour in general and in work situations in particular. Both the Danes and French agree that Danes often also use “irony”, not least “self-irony”, at their work place, which seems incongruent to the French. They also agree that the French – when using humour at all in work situations – prefer a *play on words*.

Some quotations will illustrate these differences. For example, the Danish director of a Danish enterprise in France refrained from using humour with his French colleagues for the following reasons<sup>2</sup>:

I guess I would be careful in using humour – towards a French person – with whom I am not acquainted beyond professional relations [...] I do not think it would be correct (to use humour in professional relations). For – it is also *important that you are perceived of as serious* [...] (my emphasis)

Another Dane, a woman in her fifties, had a similar statement about using humour and irony with French people:

---

<sup>1</sup> The interviews were semi-structured around six introductory questions, which were in the natural course of the interview followed up by probing and specifying questions (Kvale & Brinkmann 2009).

<sup>2</sup> Quotations from the interviews are translated word to word from Danish or French.

you can use humour with someone [...] who has been in contact with Danes for a long time — for instance, the man from whom I order travel, him I can call up and say “*Oui, bonjour, c’est la vieille*”<sup>3</sup>, [...] and then he calls me up and says “*Salut la vieille*”<sup>4</sup>. But --- it is not something, which does harm him [...] it is more self-- self ironical

This is an example of *self-deprecating* humour, the legendary Danish self-irony, used here to transcend the social taboo of not talking about a woman’s age. The irony was started by the target, the “butt”, of the irony herself, and only taken up by her interlocutor because he had known her professionally for a long time. It is difficult to imagine the French man suggesting this act of humour by himself, on his own initiative. The Danish interviewee explained the problem in these terms:

You have to be very careful about humour, because we have a *humour based incredibly much on irony*, and that the French do not understand [...] we also have a form of humour where we sort of laugh at ourselves [...], that the French are incapable of [...] [Q: *self-irony?*], no, they don’t have that, and we on our side easily become ironic on their behalf, [...], because they are so self-, self-, well not egoistic, but very self-asserting, in some way [Q: *self-important?*]<sup>5</sup>, yes, self-important, that we try, you know, all the time not to be [...], irony you have to be very careful about, they don’t understand it and they can get incredibly offended [...] (my emphasis)

The French interviewees had exactly the same opinion on the Danes’ use of irony as the Danes themselves, as witnessed by this young French woman:

I think Danes are very ... ironical in the sense that they can say negative or nasty things to make people laugh without there being anything profound behind the words. But in France, at least in work situations, those are a type of remark that you would never permit yourselves [...] [Q: *How did you react to the irony? Did you understand?*] I understood, but I think I understood because I lived with a Dane, so I was already acquainted with this type of humour. I was just somewhat surprised that this type of humour is also used in professional life and not only in private life. [...] I told myself that I would never do that.

Both Danish and French subjects mentioned a play on words, puns, as being typical of French humour. The statement by one of the Danish interviewees that «French humour is more plays on words [...] They are really good at it», was only corroborated by the French subjects. One of them, a woman in her mid-twenties, said:

---

<sup>3</sup> “Hello, it is the old one/the old lady”.

<sup>4</sup> “Hi, old lady”.

<sup>5</sup> In Danish ‘selvhøjtidelig’.

I love play on words [...] my father adores humour, he plays a lot on words [...] I have not mastered the [Danish] language enough to play on words. [...] Danes use a lot of irony; I only use it a little; actually I am not aware if I use irony.

In general she judged that Danes have less humour than French people and that irony is not particularly funny.

These differences in the conception of humour between the two parallel the general judgments of “the other”, as elicited by the question: “Describe the French/the Danes in three words”. The five Danish interviewees gave these characteristics of the French:

- formal, respect authorities, knowledgeable
- pride, formal, elegance
- “a ‘no’ in France is not necessarily a ‘no’”
- considerate
- respectful and gallant, they complain all the time, individualists

These features which for Danes distinguish the French concur with the French interviewees’ opinions of the Danes, whom they see as:

- open, focus on private life, relax
- cosy<sup>6</sup>, individualists, critical
- motivated, efficient, happy
- evasive, bon-vivants, intolerant/critical
- enthusiasts, intellectually honest, arrogant/criticise everybody

The data collected in the survey (Lundquist 2009) does not suffice to substantiate that these attributes actually constitute ‘cultural differences’ between Denmark and France, but they do show that the French and Danes have different views about each other in general and about each others’ use of humour in particular, and that these views in some respect mirror each other. Below, I shall situate these results first in a linguistic-cognitive framework, and second in a sociological framework.

### 3. Humour

In conformity with the views above about cultural differences in the use of humour by Danes and the French in professional situations, not one of the interviewed persons could recall an incident of a French person using humour in their common

---

<sup>6</sup> Proposed in Danish by the term ‘hyggelig’.

work situations. Two people, however, narrated episodes with Danes creating humour in context with French colleagues; one was a “successful act of humour” consisting of a pun, made inadvertently by the Danish speaker as a *lapsus linguae*; the other an “unsuccessful act of humour” with a Dane using irony towards his French colleague. The two examples will help us illustrate some of the linguistic and cognitive mechanisms at play in humour.

### 3.1 A successful act of humour: a Danish *lapsus linguae* in French

The Danish subject, a woman then in her twenties, had a job selling Danish furniture in a department store in Paris, BHV, Bazar Hotel de Ville. She tells the story (my translation word by word from Danish):

I had to reduce the prices of furniture, Danish furniture that is, and then someone passed by and asked ‘what are you doing?’, and I said “*j’en ai marre, je monte la moitié des prix, et je baise les autres*”<sup>7</sup> [...] Within two minutes, someone else came over asking ‘what are you doing?’ and within ten minutes, ten more had come over to ask ...and then I realised that I should not ‘*baise*’ them, but ‘*diminuer*’ them, and then it was no longer fun.

What we have here is a very nice *lapsus linguae*, of which the subject became aware only later, as stated in the interview:

I couldn’t tell the difference between voiced and unvoiced ‘s’<sup>8</sup>.

In fact, the speaker confounds the verbs ‘*baisser*’ [bese] (to decrease) and ‘*baiser*’ [beze] (to fuck<sup>9</sup>), a clear example of a *minimal phonetical difference* with a huge social effect, since many people, including the French boss, “came over” in order to have the ‘act of humour’ repeated.

#### 3.1.1 *The linguistic technique of the witz*

The functioning of this unconscious pun can best be described with the linguistic mechanisms of the ‘witz’, so clearly exposed by Freud in his seminal work from 1905 “*Der Witz und seine Beziehung zum Unbewussten*”/“Wit and its relation to the unconscious” (Freud 1999). Freud commences his very elegant treatment of the witz – in English sometimes called ‘wit’ and conceived of as an uncountable noun, and

<sup>7</sup> ‘I am fed up, I increase half of the prices, and I fuck the others’.

<sup>8</sup> As all consonants are unvoiced in Danish, the Danish speaker was probably eager to be over-correct, with the result of being incorrect.

<sup>9</sup> ‘*baiser*’: vulgar usage ‘to screw, lay, fuck’, Le Robert & Collins Senior.

sometimes termed by the countable ‘witticism’ or ‘joke’ – by investigating its technique. He starts out with this example:

[...] Ich sass neben Salomon Rotschild und er behandelte mich ganz wie seinesgleichen, ganz famillionär”<sup>10</sup> (Freud 1905/1981: 13).

Freud illustrates the linguistic technique of this witz in the subsequent formula, in which we see how two words are condensed or abbreviated into one:

Famili		är
<u>Mili</u>	o	<u>när</u>
Famili	o	när

Elaborating on the linguistic form of the witz, Freud describes it as an ‘*abbreviation*’ with ‘*concentrating force*’; in technical terms «a *condensation* with *substitution*» (Freud 1999) (my emphasis). Brevity is crucial for the humorous effect of the witz, but not, however, «in itself witty, otherwise every pithy remark would be a witticism. The brevity of witticisms must be of a particular kind» (Freud 1999: 19), in the sense that the «brevity of a witticism is often the result of a particular process which leaves a second trace – the formation of a substitute – in the wording of the witticism» (op.cit. 20).

What is important is the effect of “pleasure” caused by a witz, about which Freud has the following to say (and much more of course). Distinguishing between “tendentious jokes” and “innocuous jokes” (op.cit. 113), the former permit one to get around some external obstacle, whereas the latter make it possible for an internal obstacle to be overcome and an internal inhibition lifted. To this it may be added that «the gain in pleasure caused by a joke corresponds to the saving in psychical expenditure» (op.cit., 114) – «here we begin to suspect that *economizing in psychical expenditure* altogether must have a far more comprehensive meaning, and we must consider it a possibility that a closer definition of this still very unclear concept of “psychical expenditure” may bring us closer to the essential nature of the joke» (op.cit. 115; my emphasis). What interests us in particular is the notion of “pleasure” in combination “with lift of internal inhibitions and external constraints”, which seem to be called for in stressful cross-cultural and cross-linguistic work situations.

In Freudian terms, the *lapsus linguae* mentioned above functions via a very economical technique with the minimal opposition between unvoiced and voiced ‘s/z’:

---

<sup>10</sup> “[...] I sat next to Salomon Rothschild and he treated me just like his equal, quite familionairely”



baissér' [bese]  
 'baiser' [beze]

Due to its linguistic brevity, this witz is a radical 'abbreviation' with a 'concentrating force' of two contrasting domains, namely 1) reducing the price of furniture and 2) laying (female) colleagues<sup>11</sup>, or in more general terms, the concentration of 1) a buying/selling script and 2) a sexual script.

### 3.1.2 *Cognitive aspects of the witz*

The idea of two contrasting domains lays the foundation for the so-called "scriptbased semantic theory of humour" of Viktor Raskin (1985) and his followers (e.g., Attardo 1994), who describe the structure of an "Act of humour" as "the simultaneous perception of two different things", "a funny reconciliation of two irreconcilables", and more specifically as the "compatibility with two different scripts, which are opposites." (Raskin 1985: 99).

With the notion of *scripts* we approach *linguistic-cognitive* theories representing knowledge and its role in understanding, which have been around since Minsky (1968). *Scripts* later developed into concepts such as *mental models* or *representations* (Johnson-Laird 1983) and *mental spaces* (Fauconnier 1994), which all refer to the mental picture called to mind when deciphering a linguistic message. This is the case for Gilles Fauconnier's theory on *mapping between mental spaces* (Fauconnier 1999), which, linking language to cognition, also seems promising for the explanation of the witz, since "mappings between domains" are considered to be «at the heart of the unique human cognitive faculty of producing, transferring, and processing meaning» (op.cit. 1). A special case of mapping is 'blending', described in these terms:

Blending is in principle a simple operation, but in practice gives rise to myriad possibilities. It operates on two input mental spaces to yield a third space, the blend. The blend inherits partial structure from the input spaces and has *emergent structure of its own*. (Fauconnier 1999: 149. My emphasis).

Fauconnier's concept of mental space is close to Raskin's scripts but expands the cognitive dimension; what is new is the emphasis put on the *emergent structure* of the blend, that is, on the creation of *new meaning*. In the example above of 'Rotschild treated me quite familionairely', we experience the blending of the two mental spaces inherited from 'familiarily' and 'millionaire' respectively, whereas *the emergent mental space* equals the effect of the witz.

---

<sup>11</sup> The director of the department had the reputation of being a womaniser.

Within another branch of cognitive linguistics, *cognitive pragmatics*, special uses of language such as metaphor and irony have been treated within the general framework of the so-called *relevance theory* (Sperber and Wilson 1995). Cognitive relevance in general is described in ‘economic’ terms, as corresponding to *maximum cognitive effects* (in terms of inferred assumptions) with *minimal cognitive efforts*, i.e. saving psychical expenditure. This links neatly to Freud’s assumption on the one hand about the ‘brevity’ of a witz being its very essence, and to the considerable (cognitive) effect caused by a witz on the other.

A final remark about the cognitive side to the *lapsus linguae* above, is that the two ‘conciliated’ domains or scripts – the selling/buying and the sex scripts – are so general across cultures that they are also easily perceived as such in cross-linguistic and cross-cultural contexts. This is far from being always the case with scripts, some of which may be of a very ‘local’ nature, only recognisable for people with appropriate cultural background knowledge. An example is the pun *Galleries Lafaille* (“Gallery Bankruptcy”), heard on the French radio (November 2008) during the looming financial crisis. Here we have a play on words with the name of the big French warehouse *Galleries Lafayette*. Linguistically, the technique is subtle. It is very *economic*, as phonetically the linguistic substitution exists in *one minimal pair* only, [e] as opposed to [i]:

Galleries Lafayette	[fa-jet]
Galleries Lafaille	[fa-jit]

This minimal linguistic substitution, however, suffices to bring together two scripts, viz. 1) the name of a warehouse and 2) bankruptcy, of which the identification of the first requires cultural background knowledge. The blending of the two semantic scripts creates *the new meaning* in the sense of “Galleries Lafayette is to collapse”, which is, of course, far from funny in itself.

### 3.1.2 *Analysing the lapsus linguae*

The *lapsus linguae* reported above (‘je baise les autres’) presents, in cognitive terms «a funny reconciliation of two irreconcilable scripts» (Raskin 1985), a «blending of two mental spaces with an emergent structure of its own» (Fauconnier 1999), an emergent structure which equals the effect of its witz. The witz, innocuous and unconscious as it is, transcends a sexual taboo in a pleasurable way, and thus cheats (self)control systems «in the form of [...] outer and inner censorship, [and] power relations» (Freud 1905). Saving psychical expenditure by its ‘linguistic shortcut’, it releases a lot of energy. This energy is transferred in the context into at least ten colleagues and the boss ‘coming over’ and mentally involving themselves in the ‘act of humour’, which therefore can be said to function here as a perfect ‘mediator’ and ‘creator of a new social tie’ (Latour 2005).

Another aspect of the *lapsus linguae* example above can be explained via pragmatic principles as proposed by Raskin (1985). The normal way of conveying information in ordinary communication is the *bona-fide communication mode*, that is, the “no lying, no-acting, no-joking mode” (Raskin 1985: 88). In humour, the *bona-fide mode* is changed into a *non bona-fide mode*, without the perception of which the words would simply not be taken as a joke, but as a simple, but probably strange, assertion. This goes, as we shall see, for the frequent use among Danes of irony, an example of a *non-bona-fide* communication, often not perceived as such by the French (see 3.2). In the *lapsus linguae* example ‘je baise les autres’, the language distortion is so evident that it is instantly interpreted as a *non bona-fide communication* by the native speakers, with no risk of being mistaken for a *bona-fide communication*. A lesson to be drawn from this example, with some precaution of course, is that as a foreign language speaker, one should not fear making a *lapsus linguae*, since it may function as a mediator of new social ties, and of more relaxed social ties at that.

The example also shows the possibility for the act of humour to neutralise linguistic, cultural, and professional barriers and create a shared world, also in cross-cultural and cross-linguistic situations. Or maybe even *especially* in cross-cultural and -linguistic situations, in which a speaker may inadvertently happen to make puns in the form of *lapsus linguae* without the interlocutor taking offense, because the latter is aware of the involuntary linguistic distortion. In fact, for foreign language speakers, «their ignorance is pardonable because it is foreign languages that they are not proficient in» (Raskin 1985: 185).

This tolerance towards non-native speakers is perhaps even more common for the French with their ‘respect’ for others and ‘consideration’, two characteristics noticed by the Danish interviewees. In order to explain the French preference for play on words over irony together with their general tolerance towards linguistic lapses, we shall refer to a sociological framework.

### 3.1.3 *A sociological approach to French plays on words*

In every society there are certain norms for social behaviour. In French society, these norms seem to point in the direction of civilised manners, elegant conversation and stylistic conventions. This has been explained very convincingly in the framework of the ‘historic sociology’ by the German sociologist Norbert Elias in his books “Civilizing process. History of Manners. State formation and civilization” (1994 (1939)) and “The French court society” (*Die höfische Gesellschaft*, 1969; English version 1983/2006). According to Elias, in general

The social units that we call nations differ widely in *the personality structure* of their members, in *the schemata by which the emotional life of the individual is*

*molded* under the pressure of institutionalized tradition and of the present situation (Elias 1994: 27. My emphasis).

This explains why characteristics stemming from a particular form of process of civilisation «bears witness to a particular structure of human relations, to a particular social structure, and to the corresponding form of behaviour» (Elias 1994: 47).

As far as French civilisation is concerned, cultural identity has been shaped (as for example compared to the German) by a very specific process of civilisation emanating from the king and the court. Thus, for Elias, the specific affective moulding of the French is a

direct continuation of the *courtly-aristocratic tradition of the seventeenth century* [...]. When the bourgeoisie became a nation, much of what had originally been the specific and distinctive social character of the courtly aristocracy and also of the courtly-bourgeois groups, became, in an ever-widening movement and doubtless with some modification, the *national character*. *Stylistic conventions*, the *forms of social intercourse, conversation, articulateness of language* and much else – all this is first formed in France within courtly society, then slowly changes, in a continuous diffusion, from a social into a *national character*.” (Elias 1994: 30. My emphasis).

This leads, still according to Elias, to a *hierarchisation* of society, moulded by people who «could parade their status while also observing the subtleties of social intercourse, marking their exact relation to everyone above and below them in their manner of greeting and their choice of words». In short, their social behaviour was shaped in order to “*make visible the hierarchy of society*”. This “pressure of court life” with

the necessity to distinguish oneself from others and to fight for opportunities with relatively peaceful means, through intrigue and diplomacy, enforced *a constraint on the affects*, a *self-discipline* and *self-control*, a peculiarly *courtly rationality* (Elias 1994: 268. My emphasis).

Compared to the former warrior culture, the court society thus constitutes “a marked shift from external to internal constraints”, from control over others to *control over oneself*. These civilising constraints create social norms:

built into the personality which – as in the form of good manners, norms, values, ideas or a good conscience – form a highly valued component of the ‘self’ and of self-respect that is both personally and socially indispensable. [...] Individuals who did not compete or lost their self-control and acted impulsively under acute affective pressure had to contend with social failure and quite often the loss of their position. (Elias 1969/2006: 239-41).

In his wish to state «common structural features of social movements and developments in paradigmatic form» (Elias 1969/2006: 232), Elias introduced the notion of *figuration of personal interdependencies*, the specific form of which characterises a social configuration at a given moment. Different figurations consist of specific *chains of interdependencies*, which are subject to constant changes. What happened with and after the special social figuration of the French court society described above was that the “chains of interdependence lengthened and tightened” (op.cit. 240), but with a continuation of themes. Thus the overall direction of the development of the courtly society went towards «increasingly interdependent, larger and more complex forms of human association» (op.cit. 233), leading to a more and more centralised state:

the development of more and more centralised states with an increasing division of functions, and of larger and larger royal courts or, at a later stage, larger and more comprehensive centres of government and administration, the growth of cities, increasing monetarisation, commercialisation and industrialisation – all these are merely different aspects of the same overall transformation. (Op. Cit., 233).

Under Louis XVI, the court society of the *ancien régime* – this «figuration with a multipolar balance of tensions» (op.cit. 288) – had become a petrified figuration «of ruling elites [...] caught by their opposed tensions as in a trap.» (op.cit. 291-92). This transformation in interdependencies can contribute to an understanding of «the end of the *ancien régime*, of the adoption of violence by groups previously excluded from control of the monopoly of force and from the power that went with this control» (op.cit. 289).

However, the *hierarchical social organisation* and *elitist thinking* still permeate French society, if one is to believe recent studies on, for example, the school and education system in France (Milner 1984/2009) and on the organisation and management of firms and work in that country (d'Iribarne 1998). Several organisational studies point to the French *hierarchisation* with its “empowerment from the top” (d'Iribarne 1998), its hierarchy of authority with «a preference for a *centralized administration* based on personal relationships, and a respect for status» (Lubatkin et al. 2005: 879). Also the ‘logic of honour’ which is said to regulate personal relations in professional contexts (d'Iribarne 1989: 27) can be seen as a continuation of structural features typical of the figuration of personal interdependencies in the court society. In fact, typical ‘frenchness’, ‘l'étrangeté française’, is seen by d'Iribarne as identifying with characteristics such as dignity and ‘noblesse’, attributes referring here to personal character and not to a social class. The emphasis on *distinctiveness* as opposed to *commonness* still constitutes the make-up of the French people and of France, this *society of ranks* (d'Iribarne 2006).

Let us summarise what can be deduced from the quotations above about the collective programming of the mind of the French. Personal interdependencies seem, in France, to be founded on respect and self-respect, dignity and noblesse; on control of affect, on hierarchy and rivalry aiming to maintain one's status and honour, all features which, as we shall see below, differ radically from what has come out of the civilisation process in Denmark, moulding quite another kind of mental programming. In section 4, we shall see how the preference for different kinds of humour between the French and Danes, at least when it comes to professional settings, corresponds to the social norms and constraints shaped by the two different processes of civilisation.

### 3.2 An unsuccessful act of humour: Danish irony in a *lingua franca*

The following episode told by one of the French interviewees leaves no doubt that Danes' use of irony might be perceived as a lack of respect, causing considerable frustration and nuisance for the French interlocutor, moulded by another kind of civilisation process.

The interviewee, a French man in his forties, who had lived in Denmark for seven years with his French wife, had experienced severe problems in his interaction with Danes because of their use of irony:

I tend to find Danish humour very often repetitive, so that after a certain time, I end up asking myself, hell, is this a joke or is it serious? [...] I feel a little lost and then, as a consequence, I am not really sure whether we are in the domain of humour or whether we are out of it<sup>12</sup>.

On his own initiative, he told of an incident when one of his Danish colleagues met him one morning at 8.15 a.m. at the office with the words translated here from French (used in the interview), but originally presented in English which was the common language, the *lingua franca*, at the office:

In Denmark people start working at 8 o'clock – *although there was no one in the museum, we were in practice the only two people*<sup>13</sup> – in Denmark you have to meet at 8 o'clock, since, you see, we are in Denmark and we work from early on.

---

<sup>12</sup> Translated from French.

<sup>13</sup> This comment from the interviewee clearly shows the bringing together of two opposing scripts: 1) everybody meets at 8, and 2) at 8.15 there is no one there.

First the French person took it as a joke, but

it came up so many times with the same person that after a year, I began getting there at 7.45 [...] Actually, I did not know any longer if it was humour since it was repeated a little too regularly.

Asked whether he recalled another incident, he told the following story, which took place shortly after his arrival in Denmark, when he had trouble getting into evening courses to learn Danish. A colleague of his met him with these words (translated from French, which again was the rendering of the Dane's wording in English):

Anyway, in order to learn Danish, well, there is no choice, you have to divorce and marry a Dane/a Danish woman<sup>14</sup>.

This remark was taken as a severe offence, explained by the French person in these words (translated from French):

It came up very coldly in a discussion, and as it was at a moment when I was trying desperately to learn Danish, it was very hurtful. The big, big danger is that, as you do not understand, because there is a *cultural difference*, you get paranoid very fast, and yes, I think that I have become paranoid. Now, if somebody makes a joke, I have a tendency to interpret it *first as not being a joke*, and then I start to think about it, well, don't exaggerate, *it was probably a case of humour*. (my emphasis)

### 3.2.1 Cognitive explanation

This explanation is noteworthy, since it shows that the linguistic shortcut and the saving in psychical expenditure together with a release of energy did *not* take place, because the listener hesitated as to whether to interpret the words as a *bona-fide* or as a *non-bona-fide* communication. His interpretation took place in several steps, of which only the last, on second thoughts, went in the direction of a *non bona-fide* communication, i.e. of humour, which had of course evaporated by then. Thus, what the Danish speaker probably intended as humour was certainly not received as such by the French interlocutor, an interpretation problem inherent in irony for several reasons.

Described via scripts (cfr. Raskin above) or of blending of mental spaces (cfr. Fauconnier above), irony consists of presenting a script or a mental space in two forms, an (explicitly) asserted and a (an implicitly) negated form:

Script A +    Script A –

---

<sup>14</sup> Told in French: «De toutes façons, pour apprendre le danois, écoute, il n'y a pas le choix, il faut que tu divorces et épouses une Danoise».

Only if the asserted form is interpreted as its own negative counterpart does the act of humour succeed, which is far from being always the case, as seen above. Instead of *pleasure*, an act of irony risks creating *discomfort* and *paranoia*. Still, it is a *mediator*, forming new social ties, but this time of a negative kind.

In the conclusion I shall suggest an explanation for the Danes' frequent use of irony, but let us first listen to the Frenchman's own explanation. He imputed his paranoia about the Danes' use of irony to an "incompréhension culturelle" and different "bases culturelles", but he also gave an explanation in linguistic or phonetical terms:

I understand it as being humour probably only 50% of the time, because it is said with the same intonation. [...] In French, we have a lot of intonation, we have a tendency to stress if it is a joke. In that way at least you see it coming. In Danish, the jokes arrive flatly in the discussion. *There are no forewarning signs*. Once it is out, I need some time to reflect whether it's a joke or what. (my emphasis).

I shall propose a more elaborate explanation in linguistic and social terms below.

### 3.2.2 Linguistic explanation

The use of a *lingua franca* may explain why the French interlocutor did not perceive any linguistic 'warning' signs in the messages he only on second thoughts interpreted as being ironic, without however ceasing to be offended: either the Danish sender did not know how to express these intentions in the English *lingua franca*, or the French receiver did not spot them. What is almost sure however, is that had the Danish speaker used his native tongue, the Danish language would have offered an impressive list of warning signs in the form of the so-called modal particles (*jo, da, så, bare, nok, vel, hellere*, etc.) which are so frequently used in Danish. Thus it is reasonable to suppose that the remark "Anyway, in order to learn Danish, well, there is no choice, you have to divorce and marry a Danish woman" put in Danish would have contained a set of such 'bonding' particles, which would have created a context of presupposed intimacy, signaling that the remark should be taken as a *non bona-fide* communication containing no threat, neither personal nor social. The unsuccessful act of irony of the utterance would probably have been changed into a successful one by adding one or more of these particles characteristic of the Danish language:

*Så må du jo hellere lade dig skille og gifte dig med en dansker*<sup>15</sup>  
*Så må du vel hellere lade dig skille og gifte dig med en dansker*  
*Så må du nok hellere lade dig skille og gifte dig med en dansker*

---

<sup>15</sup> These adverbials are close to impossible to translate.



Så må du *jo bare* lade dig skille og gifte dig med en dansker  
 Så må du *vel bare* lade dig skille og gifte dig med en dansker  
 Så må du *nok bare* lade dig skille og gifte dig med en dansker

Så må du *jo nok hellere bare* lade dig skille og gifte dig med en dansker  
 'Then you had *better/supposedly better* divorce and marry a Dane

All languages have such *pragmatic particles*, including French, as in, for example, *enfin, alors, finalement, je pense, vous savez, si vous voulez*, etc. The characteristic of the Danish particles, however, is that they are short and they consist of condensed linguistic material, which explains why they are widely used. Furthermore, the short form together with an often flat intonation makes these bonding particles less audible to a foreign ear, for which reason they may easily escape foreign interlocutors. These modal particles are so prevalent in the Danish language that it has been characterised as a “receiver-oriented language”, whereas the French language belongs to “world-oriented languages” (Durst-Andersen 2001; Durst-Andersen & Lange 2010).

In the reformulations above, the bonding particles introduce *different voices*, one of which asserts the content of the sentence, while the other negates it<sup>16</sup>; one which signals a *bona-fide* communication, the other a *non bona-fide*. In linguistic terms, irony is an instance of ‘polysemy’, i.e. the presence of several voices (Ducrot 1984<sup>17</sup>). A polyphonic approach to irony is also adapted in Sperber and Wilson, who see it as a figure of “echoing”; ironic utterances are echoing since they are “second-degree interpretations of somebody else’s thought (with a certain sceptical, amused, reproving attitude)” (Sperber and Wilson 1995: 238).

It would certainly be of interest to ask why Danes show such a penchant for speaking with two voices, beyond the fact that their language presents easy access to polyphony. What is the point of stating and negating a state of affairs at the same time, of blurring the borders between *bona-fide* and *non bona-fide* communication? The best solution I can come up with for now is the description ‘evasive’, as proposed by one of the French interviewees, which may be due to the fact that social norms in Denmark in general enforce a *conflict avoiding* and *consensus seeking behaviour*, which also prevails in professional settings. The specific historical and social process by which the Danes have been civilised may shed light on this particular national character.

---

<sup>16</sup> For a detailed description see Krylova (2007).

<sup>17</sup> For polyphony in Danish see Durst-Andersen (2007).

### 3.2.3 *A sociological approach to the Danes' use of irony*

Though Elias himself has offered no description of how the Danish society was moulded by a specific process of civilisation, Danish historians and sociologists have been inspired by his works, often starting, however, their description of the Danish civilising process at a later moment. The defeat of the “double monarchy” in the Second Schleswigian War of 1864 with the ensuing loss of the duchies to Germany reduced Denmark from a big composite state to a “Big Small State” (Østergaard 2006). Hence, «In Denmark [...], national identity took the form of *the reactive defensive nationalism of a small state*, not the aggressiveness of a large and strong state» (Østergaard 2006: 83. My emphasis). Danish identity is conceived primarily in terms of language and culture, enforcing a movement towards an *inner coherence* helped by an exact congruence of state, nation, language and people<sup>18</sup>. The moulding of the Danish identity, as seen by Østergaard, thus ideologically

differs from French thinking, in which the state-nation is a core concept, and state and nation mutually help in defining the other. The ethnocultural notions of *Volksgemeinschaft* (Danish *Folkefællesskab*) as an organic, linguistic or racial community were first formulated in the early nineteenth century by German intellectuals who sought to distance themselves from what they saw as the shallow rationalism and cosmopolitanism of the Enlightenment and the French Revolution. (Østergaard 2006: 83).

The movement of the Danish *Folkefællesskab* was fuelled by the Danish priest and philosopher Grundtvig and his idea of ‘folkelighed’ (‘popularity’), «with the responsibility towards the whole and the obligation to involve all the members of the national unit» (Jespersen 2004: 212). Grundtvig and his followers paved the way from absolutism to democracy and to the creation of a *particular Danishness*, the very icon of which is the modern *Danish welfare state*:

the modern Danishness and the way in which Danes perceive themselves is especially constituted by the norms, values and rules of behaviour of the welfare system. (Jespersen 2004: 213).

The identification of Danishness with the Danish welfare state can, however, be traced back further to the Danish *ethnie* of the middle ages as formed in village communities in an agrarian society «marked by shortage and economic and social distress» (Jespersen 2004: 213). Jespersen sees a direct link between the *ancient*

---

<sup>18</sup> The following words penned in 1872 by the Danish poet H.P. Holst have «become something of a motto representing the mood of the time» (Jespersen 2004: 201): «Every loss can find recompense again! Each outward loss must turn to inner gain!».

*village community*, Grundtvig and «the Danish welfare model which in the course of time has become synonymous with Danishness»:

However, it was first with the impact of Grundtvig, towards the middle of the nineteenth century, when he transformed the old rules of conduct and norms of behaviour from the old agrarian society into a coherent national utopia, that *the values of the ancient peasant society were promoted to common Danish values*, and thus to an important element of the identity – further defined and developed by politicians, artists and writers from the era of Romanticism onwards – which pointed the way forward to the modern welfare society. (Jespersen 2004: 21. My emphasis).

Jespersen agrees with Mellon (1992) that Danishness has a “tribal” character<sup>19</sup>, described as the Danes’ *camp-fire mentality*:

The particularly Danish solidarity based on a tribal feeling [...] could perhaps better be compared to a camp fire. All the members of the company assemble shoulder to shoulder around it, with their backs to the darkness outside the circle of light from the fire, and *warm themselves with feelings of security and comradeship* and the heat form the large communal bonfire. [...] There is no sense of a ‘us and you’ feeling, but of an *introverted ‘us-us’ mentality*. (Jespersen 2004: 209. My emphasis).

To these general aspects of the social disciplining of Danes into what they are now, should be added one legal and one political aspect. First, The Danish Law in 1683

legally enshrined *the Danish model of contracts*, that is that the state would not interfere in agreements between two parties and that a verbal contract was just as binding as one in writing. This survival of the way in which things were done in the village collectives later also *formed the basis for the modern Danish labour market legislation*, which left as much as possible to the parties involved themselves and continues to be an important condition for calm in the labour market, and so for the efficiency of the Danish welfare model. (Jespersen 2004: 212. My emphasis).

This contractual aspect of the Danish way of social organisation, which was the only way for Denmark to survive as a *survival unit* (Elias referred to in Kaspersen 2008), has also been seen as typical for the political tradition of *national consensus*; underlying the conflict in the late 19<sup>th</sup> and early 20<sup>th</sup> centuries between right and left parties

there existed a common project that generated some level of consensus – the national project [and a] series of social acts introduced as of 1891 inaugurated a new

---

<sup>19</sup> «The Danes are not a nation ... they are a tribe, this is the strength of their fellowship and the reason that they have unshakeable trust in each other», Sir James Mellon quoted in Jespersen (2004: 6).

period in Danish history. For the first time the state budget contained an account of social expenses next to the expenses of the King, the Church, the army and the navy (Kaspersen 2006: 112).

And in 1915, even the old party of the right, Højre, having changed its name to the Danish People's Conservatory Party,

introduced a socio-political programme as part of its working basis that in most respects followed the line of the victorious concept of *the equivalence of social responsibility and Danishness*. Thus after decades of bitter struggle and heated debate, a *national consensus* as finally formed which paved the way for the project which was to become the most important rallying point of Danishness and its most striking manifestation in the twentieth century – *the all-inclusive and all-embracing welfare state*.” (Jespersen 2004: 206-07. My emphasis).

The Danish conflict-avoiding, consensus-seeking welfare state is presently under pressure from «the rapidly increasing immigration and its transforming of the previously homogeneous Danish nation», on the one hand, and the

constantly growing pressure for integration from the EU, in the form of demands for harmonisation of laws, taxes and duties [which] based on Roman Law, collide head on with the traditions behind the Danish model of contract, which, as often mentioned, is the very foundation of the welfare society (Jespersen 2004: 215)

on the other. Although these pressures from the outside world on the Danish *campfire identity* are somewhat turning the *us-us* mentality into an *us-them* mentality where caustic sarcasm might replace good-natured humour and irony, I shall consider the *campfire identity* and the *us-us* mentality together with egalitarian principles as still being characteristic of the Danish collective identity; in comparison, we find the French identity which was mainly moulded by the court rationality of the 17<sup>th</sup> century and its ensuing hierarchical and elitist thinking.

The Danish tradition for egalitarian organisation is continued both in the school and higher education system with its principle of *unity* (Korsgaard 2006), and in organisations and management. Indeed, ‘Scandinavian management style and culture’ has been characterised and valued because of its flat structure, «low power distance, egalitarianism, collective responsibility, and corporation» with consensus decision-making (Lubat 2005: 875).

In the following section, I shall try to explain the differences in the use of humour, and especially of irony, in the light of these different ‘national cultures’, the results of two radically different processes of social disciplining.

#### 4. Humour and national identity

How do the differences between the uses of humour fit into these two radically different processes of social disciplining described above? In this section I shall try to show some evident relationships linking the preferred types of humour in professional settings – the play on words in France and irony in Denmark – to the different national identities in the two countries, which as we saw above, were shaped by two radically different processes of civilisation: in France, spreading and descending from the court society; in Denmark pervading and ascending from the village communities.

The play on words fits perfectly into a society where courtly manners in the form of conversational norms for affect control, for elegance, respect and distinctiveness reign. Irony would not only risk being perceived as ‘common’, low and vulgar, as lack of self-control and an offence to good manners, but also, if it takes the form of self-irony, put at risk one’s own status and position in the community of ranks, as also presented by the professional hierarchy. In such a context, it is no wonder that irony will instantly be perceived as being not a non bona-fide act of communication, but as a bona-fide act, and a threatening act at that. In contrast, a play on words is directed neither at the interlocutor in the form of the 2<sup>nd</sup> person, nor at the self in the form of 1<sup>st</sup> person, but at language itself, i.e. a 3<sup>rd</sup> ‘person’. The play on words is perceived as an act of verbal and social distinctiveness<sup>20</sup>, without any inherent threat either to oneself or the other.

Turning to Danish society and its process of civilisation, we find another picture. The camp-fire mentality, in which Danes seem to “*warm themselves with feelings of security and comradeship*” (see above), and which generates an atmosphere of an “*introverted ‘us-us’ mentality*”, provides a social context secure enough for irony and self-irony to thrive; in an egalitarian context, there is no imminent social risk in lowering oneself or the other as everyone is disciplined towards not rising above the others<sup>21</sup> Witness an observation made by one of the French interviewees about the cultural difference between the Danes and the French:

In France, we love to show ourselves to our advantage. In Denmark, that does not work. This entails that the situation is inversed. Danes show a tendency to put the person in front of you in an inferior position, because you cannot put yourself in a superior position. [...] In France, we love to increase our own standing, it is really part of the culture, and nobody takes it seriously

---

<sup>20</sup> It is also treated substantially in French linguistics, stylistics and rhetoric, see e.g. the impressive taxonomy of types of ‘jeux de mots’ in Guiraud (1976: 8).

<sup>21</sup> According to the (in)famous Danish “Law of Jante”/“Janteloven”.

It comes as no surprise that Danish irony was not successfully received by the French interlocutor, raised and ‘disciplined’ in a radically different process of socialisation.

In the last section I shall address the question as to whether there could be a systematic relationship between use and interpretation of humour, identity and language type.

## 5. Humour, identity and language type

In this section, I shall briefly address the old question about the relationship between language and national identity, as has been formulated in linguistic determinism (Humboldt 1822/2000; Whorf 1956), or in a total rejection of it (Pinker 2007). I shall situate this issue in the general vein of Elias’ historic sociologic approach to civilisation, and more specifically in the context of his “Symbol Theory” (1989/1991). It seems to me that it would be useful to consider whether his concept of *figuration of interdependencies* as characterising a given unit of survival at a given moment has any relation to the language spoken in that unit. In a first step, we can state with Elias that language is a social fact:

Language can serve as *the prototypical model of a social fact*. It presupposes the existence not only of one actor, but of a group of two or more co-acting people. (Elias 1989/1991: 21. My emphasis).

In a second step, it is also a fact that language, in its generic as well as specific conception, is made up of a figuration of *symbolic interdependencies*, i.e., by a system of sense-making structures and relations. These are found at all linguistic levels from lexicalisation of semantic features and up, encompassing word formation, grammar and syntax, and discursive levels of presenting and linking of information in texts (Lundquist 2006).

And finally, in a third step, we agree with Elias that

(a) people’s language itself is a symbolic representation of the world as the members of that society have learned to experience it during the sequence of their changing fortunes. At the same time a people’s language affects their perception and thus also their fortune. (Elias 1989/1991: 61).

In this context, it is tempting to address the differences in the use of and reaction to humour in professional settings observed between Danes and the French in linguistic terms also: could there be a relationship between the Danish language and the Danes’ extended use of irony? And a relationship between the French language and

the preference of the French for a play on words? At the present stage, unable to come up with a satisfying answer to that question, I shall content myself with presenting some suppositions as to such potential relations.

As far as the Danish language is concerned, we have seen that one of its basic characteristics is its *receiver-oriented* character as noticed in the extended use of pragmatic modalities; in fact, the Danish *bonding particles* facilitate the creation of an *us-us feeling*, straightforwardly inviting in other voices which are not seen as threatening in the general *camp-fire* ambience. There seems to be if not a direct, at least a straightforward link between bonding particles and the use of irony.

For the French language, the preferred form of humour, play on words, functions via other types of subtle relations between words and sense. Taxonomies of forms of *jeux de mots*, such as those found in Guiraud (1976), reveal a delicate play on different types of sound patterns, to which Saussure's model of 'associative relations' (Saussure 1916/1983: 173), created via, for example, word roots, pre- and suffixes, acoustic images<sup>22</sup>, etc., adds an interesting perspective (Lundquist in press).

On an even more abstract and speculative level, it would be interesting to discern whether the basic differences between the Danish and the French languages, as observed in the *TypoLex* project's distinction between endocentric Germanic languages and exocentric Romance languages (see Herslund, this volume), could account for the differences in the use and interpretation of humour in the two language communities. Could it be that an endocentric language with a preference for specified relations, as expressed via the verbal centre of the sentence, between nominal arguments expressing mainly the semantic feature FUNCTION, articulates a more concrete and pragmatic relation between language and its users than an exocentric language with a tendency for general, abstract relations between nominal arguments often coded for CONFIGURATION, which may be more apt for conveying distant, 'courteous' relations? Whereas it would be untimely to conclude that language – *in casu* Danish and French – has formed and moulded society – a court society as opposed to a village society respectively – it seems plausible to say that the two types of language, the pragmatic Danish language, and the Cartesian French language (Lundquist in preparation) contribute to maintaining the specific figurations of social and personal interdependencies which have emanated from their different processes of civilisation.

---

<sup>22</sup> To which could be added the specific "sound shape" (Jakobson & Waugh 1979) of French, and its phonetic and prosodic system with syllabic prosody (Herslund 2003: 72).

## 6. Conclusion

In this study of the relation between language, cognition and identity as seen through the lens of use of humour in professional settings between Danes and the French, I have situated cognitive-linguistic observations in a broader cultural-sociological framework, as presented in the works of Norbert Elias. I should like to conclude on the same tone, in the hope that a more detailed knowledge not only about *how* people of other languages and other cultures behave, but also *why* they behave as they do, namely as a result of a specific process of civilisation, can help us to understand why, for example, “the French are formal and self-important” and “Danes are evasive and cosy”<sup>23</sup>, characteristics which much too easily – in the confrontation between these two cultures – lead to cultural stereotypes, in the sense that “the French are arrogant”, and “the Danes are bad-mannered and rude”, which in fact, by the French, is conceived exactly as arrogant also.

## Bibliography

- Attardo, S. 1994. *Linguistic theories of humor*. Berlin: Mouton de Gruyter.
- Chafe, W L. 2007. *The Importance of Not Being Earnest. The feeling behind laughter and humor*. Amsterdam/Philadelphia: John Benjamins.
- d'Iribarne, P. 1989. *La logique de l'honneur: gestion des entreprises et traditions nationales*. Paris: Seuil.
- d'Iribarne, P. 1998. *Cultures et mondialisation gérer par-delà les frontières*. Paris: Seuil.
- d'Iribarne, P. 2006. *L'Etrangeté française*. Paris: Seuil.
- d'Iribarne, P. 2008. *Penser la diversité du monde*. Paris: Seuil;
- Ducrot, O. 1984. *Le dire et le dit*. Paris: Ed. de Minuit.
- Durst-Andersen, P. 2007. Det danske sprogs mange stemmer. In R. Therkelsen, N. M. Andersen, H. Nølle (eds) *Sproglig Polyfoni. Tekster om Bachtin og ScaPoLine.*, Århus: Aarhus Universitetsforlag, 163-180.
- Durst-Andersen, P. & Lange, E. (eds) 2010. *Mentality and Thought. North, South, East and West*. Frederiksberg: Copenhagen Business School Press.
- Elias, N. 1969/2006. *The Court Society*. Dublin: University College Dublin Press.
- Elias, N. 1989/1991. *The Symbol Theory*. London: Sage.
- Elias, N. 1994. *The Civilizing Process*. Oxford: Basil Blackwell. (1939 *Über den Prozeß der Zivilisation*).
- Fauconnier, G. 1994. *Mental spaces aspects of meaning construction in natural language*. Cambridge: Cambridge University Press.
- Fauconnier, G. 1999. *Mappings in Thought and Language*. Reprint ed. Cambridge: Cambridge University Press.

---

<sup>23</sup> See section 2 above.



- Freud, S. 1905/1981. *Der Witz und seine Beziehung zum Unbewussten*. Frankfurt am Main: Fischer Taschenbuch Verlag.
- Freud, S. 1999. *Wit and its relation to the unconscious*. London: Routledge.
- Guiraud, P. 1976. *Les jeux de mots*. Paris: PUF.
- Herslund, M. 2003. Essence du langage, types linguistiques et systèmes métriques. *Language and Culture. Copenhagen Studies in Language* 29: 71-82.
- Herslund, M. This volume. Predicati e sostantivi complessi: complementarità e isomorfia
- Hofstede, G. & Usunier, J.C. 2003. Hofstede's Dimensions of Culture and their Influence on International Business Negotiations. In N.G. Pervez & J.-C. Usunier (eds), *International business negotiation*,. Oxford: Elsevier Science, 137-154.
- Humboldt, W. von. 1822/2000. *Ueber den Nationalcharakter der Sprachen. Edition bilingue*. Paris: Seuil.
- Jakobson, R. & Waugh, L.R. 1979. *The Sound Shape of Language*. Hassocks: Harvester press, 1979.
- James, W. 1884. What is an Emotion? *Mind* 9: 188-205.
- Jespersen, K.J.V. 2004. *A History of Denmark*. New York: Palgrave.
- Johnson-Laird, P.N. 1983. *Mental Models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, Cambridge university Press.
- Kaspersen, L.B. 2006. The Formation and Development of the Welfare State. In J.L. Campbell, J.A. Hall & O.K. Pedersen (eds), *National Identity and the Varieties of Capitalism. The Danish Experience*. Québec: McGill-Queen's University Press, 99-132.
- Kaspersen, L.B. 2008. *Danmark i verden*. København: Hans Reitzels forlag.
- Korsgaard, O. 2006. The Danish Way to Establish the Nation in the Hearts of the People. In J.L. Campbell, J.A. Hall & O.K. Pedersen (eds), *National Identity and the Varieties of Capitalism*. Québec: McGill-Queen's University Press, 133-158.
- Krylova, E. 2007. Epistemisk polyfoni på dansk. In R. Therkelsen, N. Møller Andersen, H. Nølke (eds), *Sproglig polyfoni. Tekster om Bachtin og ScaPoLine*. Århus: Aarhus Universitetsforlag, 181-203.
- Kvale, S. & Brinkmann, S. 2009. *InterViews Learning the craft of qualitative research interviewing*. 2nd ed. London: Sage Publications Inc.
- Latour, B. 2005. *Reassembling the Social. An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.
- Lubatin, M.H., Lane, P.J., Collin, S-O. & Very, P. 2005. Origins of corporate governance in the USA, Sweden and France. *Organization Studies*, Vol. 26(No. 6), 867-888.
- Lundquist, L. 2006. Lexical Anaphors, Information Packing, and Grammaticalisation of Textual Relations. In H. Nølke et al. (eds), *Grammatica : Festschrift in honour of Michael Herslund : Hommage à Michael Herslund*. New York: Lang, 311-323.
- Lundquist, L. 2009. Humour as a mediator in cross-cultural professional settings. Examples from Danish and French. *CBP Working Papers* 66. Copenhagen: Copenhagen Business School.
- Lundquist, L. In press. *Humor og ironi i tværsproglige og tværkulturelle møder. Bona-fide, ikke-bona-fide og bona-fide-cum-humor sproghandlinger i dansk og fransk*. Skandinavisk Skriftserie.

- Lundquist, L. In preparation. Cartesianism, Pragmatism. Nouns and Verbs in French and Danish.
- Milner, J.-C. 1984/2009. *De l'école*. Paris: Verdier.
- Mellon, J. 1992. *Og gamle Danmark ... En beskrivelse af Danmark i det herrens år 1992*. Århus.
- Minsky, M. 1968. *Semantic information processing*. Cambridge, Mass.
- Pinker, S. 2007. *The Stuff of Thought. Language as a Window into Human Nature*. New York: Viking Adult.
- Raskin, V. 1985. *Semantic mechanisms of humor*. Dordrecht: Reidel.
- Saussure, F. de. 1983. *Cours de linguistique générale*. Paris: Payot.
- Sperber, D. & Wilson, D. 1995. *Relevance. Communication and Cognition*. 2nd ed. Oxford: Blackwell.
- Whorf, B.L. 1956. *Language, Thought and Reality*. Massachusetts: MIT Press.
- Østergård, U. 2006. Denmark: A Big Small State. In J.L. Campbell, U.A. Hall & O.K. Pedersen (eds), *National Identity and the Varieties of Capitalism. The Danish Experience*. Québec: McGill-Queen's University Press, 51-98.

LEXICAL SEMANTICS NEGOTIATED.  
ASSESSING THE FAIRNESS AND MISLEADING POTENTIAL OF  
ESTABLISHED AND NOVEL FOOD NAMES

Viktor Smith  
Copenhagen Business School

**1. Background and context**

A key assumption of the Typolex Research Group at Copenhagen Business School (CBS) is that it is the lexical structure of any given language that determines its typological features also on the grammatical, syntactic and textual levels – with endocentric and exocentric languages as the two archetypical extremes, at least within the Indo-European languages – and that these differences, in turn, impose different “world views” on the speakers who have these languages as their mother tongue (see the papers by Herslund, I. Korzen and H. Korzen in this volume). All things being equal, the lexicon is thus conceived as a “monolith” structure which is relatively stable across time and speakers within the same language community while the interesting variations transpire only when different languages and language communities are compared. This would immediately seem to contradict another widespread view, or rather, a plain observation: That the lexicon is the most flexible part of any language that constantly strives to keep itself in pace with the development of society, with thousands of new word being coined every year while the meaning of existing words is subject to permanent evolution and change, as opposed to the much slower pace of development of the surrounding grammar and syntax.

However, the contradiction is easily reconciled if one further specifies which aspects of the lexicon are to be scrutinized: The former view neatly applies to the level of *primary lexicalizations*, i.e. the permanent stock of free morphemes (word roots) available in any given language and their distribution across semantic domains, and to the archetypical patterns according to which new, more complex words, *secondary lexicalizations*, are coined on that basis. Here clear and robust difference can be observed between e.g. an endocentric language like German and

exocentric languages like Italian, French, and Spanish, as illustrated by Herslund (this volume). Yet on the level of secondary lexicalizations – which constitute the vast majority of words in any language of present interest – we find all the flexibility and negotiability presupposed by the latter view, within the limits given (see Smith 2000 for further discussion and examples).

This article focuses on the dynamic and negotiable aspects of the lexicon, and hence mainly on the level of secondary lexicalizations. Specifically, we address the semantic domain of *food products* where the tendencies just mentioned are highly pronounced and relate directly to real-world public concerns as to whether or not the names given to commercial food products are likely to *mislead* consumers about the food inside the package. We rely, primarily, on the findings of another ongoing CBS project, the FairSpeak Project ([www.fairspeak.org](http://www.fairspeak.org)). Relating these findings to the Typolex agenda has two major reasons. First, identifying the exact meaning of a food name (and thus also its potential misleadingness) poses several new and theoretically interesting challenges related to the process of lexicalization, even from a monolingual perspective, in terms of identifying the semantic variables in play and providing a coherent analysis of their interaction. Second, while the initial focus of the FairSpeak Project was on Denmark (and Danish), it is presently being extended to other EU countries (and languages) in view of the still increasing integration of EU food markets and the harmonization of national rules and practices of the member states in the present field. Here, the distinction between endocentric and exocentric languages (e.g. Danish or German vs. Italian or French) seems to offer a good basis for identifying the additional typological factors in need of consideration when assessing the misleading potential of specific food naming solution across the corresponding markets. However, any further synergies along these lines presuppose a clear understanding of the complexity of the issue, even when viewed from a monolingual perspective, and this is the main goal pursued here, focussing on Danish.

## 2. Materials, aims, and scope

It is widely recognized that “having a word for it” is basic to our day-to-day (re)identification and categorization of objects and events in the infinite variety of extra-linguistic reality (e.g. Gumperz & Levinson 1996; Piaget 1926; Vygotsky 1962 [1934]). In the case of food products, the name is not only crucial to the product’s cognitive identity, but also to its legal identity. According to the EU Labelling Directive 2000/13/EC, any food product sold in the EU shall carry a name, so that consumers can check if what they are buying is e.g. *cannelloni*, *spaghetti*, or *potato gnocchi*. Nevertheless, ordinary consumers as well as competitors and the

food authorities sometimes disagree with the name chosen by a manufacturer or retailer for a particular product to the extent where they take formal action against it.

In an in-depth quantitative and qualitative review of 821 Danish administrative cases on misleading food naming and labelling carried out by the FairSpeak Group (V. Smith et al. 2009; see also Møgelvang-Hansen, in press), 272 out of 1000 instances of allegedly misleading labelling elements (27,2%) concerned the product name. While the general rules and procedures for handling such cases are given by law<sup>1</sup>, the specifics addressed by complainants, companies, and authorities in their opposing assessments of individual cases are lingua-cognitive rather than legal by nature (see Smith in press for further discussion).

Thus, determining whether a food name is misleading or not is, in essence, a matter of determining whether the implicit identity statement “This is *bacon*, *apple juice*, *butter cookies*, *cannelloni*, etc.” is true or not. In turn, that is a matter of determining what exactly these words *mean* in the language in question. In some cases, the matter would seem to be settled *a priori* by national or transnational food standards containing legal definitions, e.g. for fruit juices or for chocolate<sup>2</sup>. While the legal conclusion in such cases is clear, it may be questioned whether such definitions reflect the actual expectations of ordinary consumers (Ohm Søndergaard & Selsøe Sørensen 2008). However that may be, the vast majority of food names found in Denmark and other EU countries are not legally defined. So in these cases, the question of whether the name is used correctly or in a potentially misleading way is a question of what it means as an element of the general language-system in question.

In this article, the types of arguments and common-sense reasoning put forward by the parties and authorities in the case material will be illustrated by typical examples and then transposed into more exact theoretical terms, drawing on relevant insights and empirical evidence from outside the legal sphere, mainly cognitive linguistics and experimental psycholinguistics. In this way, a basis is established for utilizing the practical experience accumulated in the case material for addressing questions of more general interest to the study of lexical semantics and the complexity of online word comprehension, lexicalization, and acquisition processes. This includes isolating the variables needed for ultimately putting the common-sense

---

<sup>1</sup> Apart from those cases where the use of a given food name is regulated by a food standard (see below), the general EU provisions against misleading food naming and labelling presently constitute the main legal source, as harmonized by Article 16 of the EU Food Regulation (2002/178/EC) and further specified in some respects by the Labelling Directive (2000/13/EC). However, applying these rules to actual cases mostly presupposes highly individualized common-sense judgments regarding the likeliness that “the average consumer” would in fact be misled.

<sup>2</sup> Directive 2001/112/EC relating to fruit juices and certain similar products and Directive 2000/36/EC relating to cocoa and chocolate products.

assumptions and predictions put forward in the cases (and in the general public debate) to experimental test. The present article presents the main elements of a conceptual framework developed for supporting such purposes. More detailed analysis of individual conflict scenarios related to food names and original experimental findings are reported in separate studies.

### 3. Two types of conflict scenarios and a fuzzy boundary

In terms of essence and the types of arguments brought forward by the parties, a basic distinction<sup>3</sup> can be made between conflict scenarios relating to:

- *Established food names* for (more or less) familiar products. Here the question is what names like *orange juice*, *coffee whitener*, or *macaroons* in fact mean to the different parties of concern (consumers, gastronomic experts, manufacturers, authorities, etc.).
- *Novel food names* for entirely new types of products with which the (average) consumer cannot be expected to be familiar. Here, the question is how names like *Halal ham* or *surimi shrimps* are likely to be interpreted when first encountered and what they may eventually come to mean to consumers, and others.

This boundary is bound to become somewhat blurred in practice for individual consumers and products, respectively, due to the steady pace of market developments. Thus, a term like *smoothie* may be familiar to a great many consumers, but novel to some. When dealing with different degrees of familiarity with a given product on the market as a whole and the naming challenges emerging from this (while allowing for some variation across individual consumers) the FairSpeak Group distinguishes between the following main categories:

---

<sup>3</sup> In some cases, the difference does not seem to be clearly recognized by the immediate actors. For example, in a case concerning a product named *crème fraiche dressing* which contained less than 20 % *crème fraiche*, the complainant treated the name as one in need for an element-by-element interpretation (which may be true of novel names, but doubtful for established ones, see sections 5 for further discussion) and insisted that the only correct interpretation was that *crème fraiche* should be the main ingredient. By contrast, the manufacturer insisted that there were similar products on the market carrying the same name with a similar content of *crème fraiche*, and that this was well known by consumers, i.e. treated the name as an established (generic) term. Case No: 2004-07-722-09899 (id 279).

- *Product repetition*, i.e. yet another brand of peach ice tea, macaroons, coffee whitener, ketchup, etc.
- *Product evolution*, e.g. low-fat version of traditionally fat cold meat product, or avocado dip containing 0,4% dried avocado powder.
- *Product innovation*, e.g. caviar “lookalike” made of sea kelp, or a replacement for “real” pizza cheese which is not primarily made of cheese.
- *Product re-incarnation*, e.g. ordinary dark chocolate marketed and labelled as a diet product.

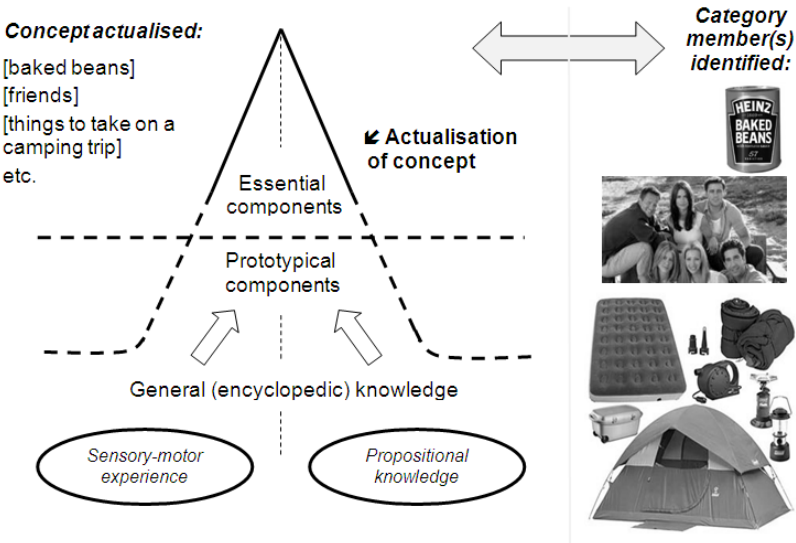
In the following, we first focus on conflict scenarios relating to food names that are treated as more or less established by all parties involved while also highlighting the variation that may arise in the essence of the conflict depending on whether the specific use of the name relates to an instance of product repetition or product evolution. Thereafter, we consider the different types of conflict scenarios that relate to food names that are comprehended as novel by all parties involved in that the product has come about as an instance of recent product innovation.

To cope with these issues, we will however first need to specify the notion of meaning a bit further.

#### 4. Some basic assumptions and prerequisites

In line with the predominant view of cognitive language theory (e.g. Evans & Green 2006; Talmy 2000), we here identify the meaning of a food name, like any other lexical expression, with a psychologically real *concept* which is conflated by language, but also serves the wider purpose of *categorization* in the course of situated thinking and acting, e.g. when shopping, eating, or developing new food products.

Following, primarily, Barsalou (1983; 1987; 1999; 2005) whose approach, in turn, incorporates earlier theorizing and experimental evidence on prototypicality and graded conceptual structure (e.g. Taylor 1989; Wierzbicka 1985; Rosch 1975; Smith, Shoben & Rips 1973), the basic “anatomy” of human concepts may be displayed as illustrated in figure 1.



*In continuation of Barsalou (1983; 1987; 1999; 2005) incorporating also cognitive theorizing on prototypicality and graded structure, e.g. Rosch (1975); Smith, Shoben & Rips (1973); Lakoff (1987); Taylor (1989).*

Figure 1. The concept as a time-bound graded structure

In the present model, concepts are not understood as static entities permanently present in the mind of anyone who “has” the concept in question (as tacitly assumed in a great many accounts), but as time-bound mental states that «originate in a highly flexible process that retrieves generic and episodic information from long-term memory to construct temporary concepts in working memory» (Barsalou 1987: 101). More plainly, concepts are seen as mental “checklists” that we compile and retrieve (actualize) for current needs from our pool of general world knowledge and experience whenever we have to distinguish some objects or phenomena in surrounding reality from others for a certain purpose. The analysis extends both to concepts actualized for mere *ad hoc* purposes – e.g. for distinguishing things that one needs for an upcoming camping trip from all the things that one doesn’t need to take – and categories that play a more permanent (i.e. steadily recurring) role in a person’s understanding of reality, e.g. baked beans or friends. There is however an important difference as to means of linguistic expression: While concepts of the former kind usually have to be rendered linguistically by more complex paraphrases, e.g. “things I need for my camping trip”, or given mere *ad hoc* names like *my camping stuff*, concepts of the latter kind will prototypically have been provided with a single, generally accepted name such as *baked beans* or *friend*. The name may well consist of two or more independent words but they will still be comprehended and used an indivisible whole (see Smith 1999/2000: 47-48, for



further discussion). In that case, we say that the concept has been *lexicalized*. As we will see in Section 5, novel food names constitute a challenging transitory case between these two extremes.

The internal structure of the concept may, in either case, be described as a complex hierarchy of mental criteria (components) used for determining whether a given real-world object qualifies as a member of the category or not. A further distinction can be made between (a) *essential components* which correspond to properties that any real-world object must possess in order to be accepted as a member of the category in question, e.g. that cheese should be made of milk, (b) *prototypical components* which correspond to properties that are a salient part of our conceptualization of the category as a whole, but do not need to be manifest with any particular exemplar, e.g. that cheese is prototypically made of cow milk, though goat or even horse milk can be used too, and (c) individual *background knowledge*, e.g. knowing that one's ex-girlfriend hates cheese. For components on all levels, a distinction may furthermore be drawn between (a) *propositional components* involving factual knowledge potentially reducible to logical propositions that are susceptible to truth-conditional evaluation, e.g. knowing which ingredients a product should contain in order to be a cheese, and (b) *sensory components* that rely on immediate recall of the taste, smell, texture, etc. of the cheeses that one has previously encountered (for further discussion, see Smith, Møgelvang-Hansen & Hyldig in press; Moskowitz et al. 2006).

## 5. Established food names: What do they mean, and who is to decide?

On this background, let us first consider some typical conflict scenarios relating to *established food names* for (more or less) familiar products, as illustrated by examples (1) - (5) below.

### (1) Whole milk or skimmed?

Product labelled *sødmælk* ≈ 'whole milk', but tastes more like *letmælk* ≈ 'semi-skimmed milk' according to consumer. Case No(s): 2004-08-274-00162 (id 272)  
CONSUMER

### (2) Mead – alcoholic, but by which means?

Consumer expects *mjød* 'mead' to be made through fermentation of honey and objects to alcohol being added later in present product. The authorities established that the product *is* indeed made through fermentation of honey, yet with even more alcohol added afterwards, a method which they find no formal grounds for questioning. Case

No(s): CONSUMER; 2003-03-274-00020 (id 85); 2003-12-274-00011 (id 196)  
CONSUMER

(3) **What's (in) a real smoothie?**

According to consumer, the product does not *taste* like what (s)he would expect of a smoothie. Complaint sustained by authorities, but the decision is motivated (only) by the *low fruit content* (0,2%). So what if the “smoothie feeling” had been simulated better? Case No(s): 2003-04-274-00297 (id 102) CONSUMER

(4) **Almonds for texture or for taste?**

*Makroner* ≈ ‘macaroons’ made of apricot kernels, not of almonds as demanded by traditional Danish recipes. Artificial almond flavour is added. Manufacturer(s) insists that this has been so since the 1940ies and that consumers like and buy the product. Case No(s): 2007-S5-274-0792 (id 735) DAF; 2007-S5-274-00795 (id 736) DAF

(5) **Nutrition and taste vs. tradition... and identity?**

Very traditional Danish meat product called *rullepølse* re-introduced in a low-fat version made of fillet and not pork belly, as demanded by all traditional recipes. Fat reduced from 25% to 3%... but has the standard recipe and *name* been violated? Case No(s): 2005-04-271-00034 (id 329) DAF; 2005-05-274-00437 (id 370) DAF

A common trait is that all parties involved seem to agree that the name has a quite distinct meaning, the question being what exactly that meaning is, or rather: should be.

In (1), the answer would seem to be given in advance in that the product is covered by a food standard<sup>4</sup>. However, the standard concerns fat content, and what the consumer is concerned about is taste. In (2), the roles are switched in that the consumer does not mention the taste at all, but objects to certain hard facts about the process of manufacturing. Yet in this case the authorities refrain from intervening for lack of any formal criteria on how the desired alcohol content is supposed to be reached in mead. In (3) there are no standards to resort to either, and, moreover, the product name (and type) may still be relatively new and negotiable to some consumers. Yet the complainant clearly seems to operate with certain fixed criteria, and these again (like in (1)) concern the sensory impression of the product. Nevertheless, hard facts in terms of low content of fruit (0,2%) end up being decisive to the outcome alone, this time in favour of the consumer. In (4), we once again observe a clash between hard facts and consumers’ (alleged) sensory preferences and liking. In addition, the number and type of evaluation criteria

---

<sup>4</sup> Council Regulation (EC) 1234/2007 establishing a common organization of agricultural markets and on specific provisions for certain agricultural products article 114, cfr. Annex XIII.

accessible to and potentially applicable by different real-life actors of relevance (manufacturers, gastronomic experts, consumers) are clearly not the same. A similar situation is displayed in (5), with the additional circumstance that the product has in fact been modified quite recently compared to most other products on the market that carry the same name. Yet the modification concerns “only” the ingredients, while the key sensory properties have been maintained. So what should be decisive?

Cases like these raise several basic questions as regards both the lingua-cognitive essence of the disputes and the optimal way of resolving or preventing them in actual food naming practices.

Transposed to the terms introduced in Section 3 above, the examples seem to indicate that different actors operate with different variants of the concepts lexicalized by the name, which display different numbers and “mixtures” of sensory and propositional components and different lines of demarcation between the essential and the prototypical ones among them. To cope with this, it is tempting to apply Putnam’s (1975) hypothesis of “division of linguistic labour” according to which members of society collaborate on knowing the exact criteria for applying a word to some particular objects, and will ultimately rely on the judgment of those members of society that have been given the status of “experts”. The question is, however, whether the expert’s final judgment will always have status as a built-in component in ordinary consumers’ variant of the concept in question, i.e. an empty slot for which only the expert may provide the right filler. This might well prove to be the case for luxury products like *caviar* and *foie gras*, but does the mechanism extend to *macaroons* or *pepperoni*? Also: Who *are* the relevant experts?

A key consideration here is also whether it is the sensory or the propositional properties of the product (and corresponding conceptual components) that are ultimately decisive for product identity (see also Smith, Møgelvang-Hansen & Hyldig in press). The examples thus demonstrate a steady “competition” between criteria that rely on (a) first-order perceptual experience with the type of food in question, e.g. the feel and taste of macaroons, and (b) second-order information available from other, mostly verbal, sources (other text on the packaging, the mass media, cookbooks, websites, etc.), e.g. about the methods and ingredients used for producing macaroons commercially. So in short: Is product identity a matter of taste or facts? And what should it be?

The latter dilemma is highlighted, in particular, in cases of product *evolution* as illustrated most clearly by example (5). Here any previous conceptualizations of the product are quite deliberately challenged by the manufacturer and may gradually change with consumers as well. The manufacturer would seem to have a good case in arguing that he is merely trying to retain the feel and taste of a popular food product while meeting consumer demands for better nutrition value. Changing the name to something completely different would most likely blur this subtle message. Instead the name is retained and the change highlighted e.g. by a so-called claim

(“Only 3% fat!”) combined with the relevant facts in the list of ingredients. It seems like a fair consideration whether different fairness criteria should apply in such cases than in those of mere product repetition (see Section 2 above). On the other hand: Should the reasoning then be extended to cases like (4) where the potential conflict is rooted more than 60 years ago and seems to have been tacitly accepted (or never questioned for lack of information?) by a great many consumers? Or even to cases like (2) where we would need to go back about 7000 years to trace the gradual evolution of meat production across time and civilizations. So maybe it is the knowledge and expectations of the “average consumer” that we should rather try to trace?

To support future best practices on such issues, ongoing FairSpeak work includes testing the limits for consumer acceptance of selected names as applied to selected products on experimental grounds by exposing different groups of experimental subjects to taste samples only (sensory product attributes), taste samples in combination with nutrition facts & ingredients list (adding propositional product attributes), and both, in combination with authoritative definitions (adding experts’ final judgments).

## **6. Novel food names: What *will* they come to mean, and how can we predict it?**

Let us now consider the quite different types of conflict scenarios that relate to food names which are treated as novel by all parties involved. In these cases, the manufacturer has taken the full consequence of having created a new product or product variant by providing it with an equally new name.

When encountering such a name for the first time, the consumers will not be able to activate any pre-determined concept whatsoever. That concept needs to be crystallized and acquired first. (S)he will therefore be left to make a situational (pragmatic) on-the-spot analysis based on the information at hand (Zlatev et al. in press) which is however bound to result in the generation of some “sketchy” ad hoc concept (e.g. «some kind of new food product which is in this package, and in other ones carrying the same name, that I’m not really sure of what is») that may later on develop into a more elaborate and permanent one.

Before addressing the conflict scenarios shown in (6) – (9) below, let us first briefly consider some overall trends in the treatment of novel food names in current administrative and legal practices, and relate these to what is known about the decoding and processing of novel versus familiar names (food names or any other) from psycholinguistic and cognitive linguistic research.

Far the most novel food names, like most other new entries in the lexicon, are created by combining existing word elements into larger units that are nevertheless

used and understood as and inseparable whole the meaning of which is not directly deducible from its parts<sup>5</sup>. In languages like Danish, English or German, the most predominant pattern is compounding (Libben & Jarema 2006; Sager 1997; Herslund, this volume), i.e. a simple juxtaposition of word stems as in *gooseberry*, *papaya milkshake*, or *butter yoghurt*, and we will concentrate on that pattern here. In current administrative and legal practices on food naming (for an overview covering the EU as a whole, see MacMaoláin 2007: 102ff), two assumptions tend to be taken for granted by authorities and complainants alike when it comes to interpreting composite names. First, that the meaning of the whole must necessarily be a function of its parts, and, second, that there is only one objective and “correct” way of interpreting the relation between the name’s constituents. That is, that *butter cookies* should objectively CONTAIN butter whereas *Alsace ham* should objectively COME FROM Alsace in France. Exceptions are reluctantly accepted for established (“generic”) names – in that no one would expect e.g. *wine gum* to contain wine or *Brussels sprouts* to come from Brussels – but for novel names judgments are less liberal.

However, existing theory and experimental evidence on the semantics and processing of complex lexical expressions call for certain modifications of these common-sense assumptions. It is thus widely recognized that “2 + 2 does not equal 4” in a compound (Ferris 1983: 66) as might be argued for full sentences (see also Benches 2006; Stekauer 2006). Thus, a compound like *snow smoothie* does not in itself tell us whether it should be interpreted as ‘white as snow’, ‘with fresh snow added’ (served as a drink), ‘to be enjoyed in the snow’ (by thirsty skiers), etc. etc. Moreover, the information deducible from the constituents themselves, even when narrowed down like above, is usually only a vague hint, an *index*, of at the full-word meaning (concept) hiding behind it (Libben 2006: 11; Smith 1999/2000, 2001: 13ff; Wüster 1959/60: 191). Thus, it takes more than just containing butter for a cookie to be recognized as and conventionally called a *butter cookie*.

As for established compounds, our familiarity with the full-word meaning may sometimes help us interpret the relation between the constituents in a sensible way, but never vice versa. Yet even that is not strictly necessary. Experimental research indicates that we do not routinely split up (decompose) familiar compounds and analyze the semantic relation between the constituents in order to retrieve their established whole-word meaning (e.g. Libben 2006; Andrews & Davis 1999;

---

<sup>5</sup> An alternative model is adding a new meaning to an existing word, which is how e.g. *caviar* has come to (also) denote a product made of artificially coloured lumpfish’s roe and not (only) “real” caviar in Danish. This is now widely known and accepted by consumers, whereas a more recent extension of the name to cod’s roe has been subject to complaint (case No: L 935-105). The fairness challenges connected with this sort of semantic shifts are however beyond the scope of this article.

Schreuder & Baayen 1995; Sandra 1990; Manelis & Tharp 1977)<sup>6</sup>. In other words, we do not need to speculate on why strawberries are called strawberries in order to understand and use the word *strawberry* correctly. Yet again, we are free to do so at any time, as shown in figure 2 (a).

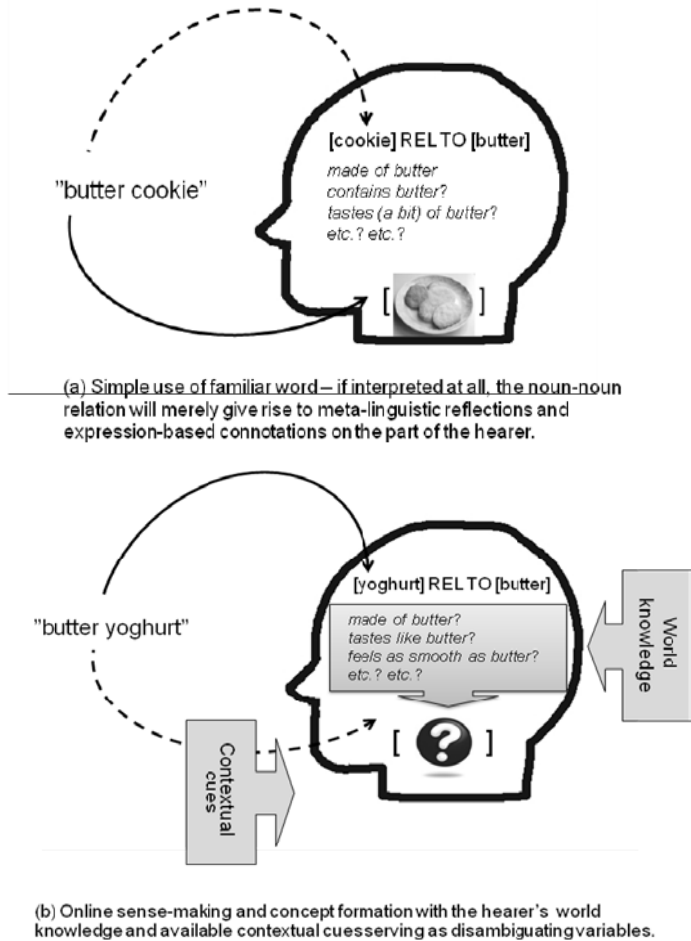


Figure 2 a/b. Processing of (a) established and (b) novel food name<sup>7</sup>

<sup>6</sup> Some studies indicate that decomposition may play a certain role for word recognition, but without reaching the semantic level. See Libben (2006), for a critical overview of existing findings.

<sup>7</sup> For supporting experimental evidence and theoretical discussion, see Libben & Jarema (2006); Benches (2006); Andrews & Davies (1999); Smith (1999/2000; 2001); Zlatev et al. (in press).

For novel compounds the case is different: Since there is no established whole-word meaning to retrieve, the consumer is bound to decompose the compound and try to make some sense of the constituents and their mutual semantic relation in their own right. When presented in isolation, the outcome is sensitive to such factors as analogies with other, familiar, compounds that share the same constituents (Krott and Nicladis 2005; Gagné 2001), and the cognitive “compatibility” of the constituent concepts (Gill & Dubé 2007; Murphy 1990), which often requires a metaphorical extension of one or both of them, e.g. *land yacht* for a luxury car (Fauconnier & Turner 2002: 357; Benches 2006). However, it has also been demonstrated that if the compound is presented in a sufficiently informative context, such default interpretations may well be abandoned in preference to an alternative one that fits the context (Gagne, Spalding & Gorrie 2005; Zlatev et al. in press). Finally, as already indicated, the need for any interpretation disappears once the relevant whole-word meaning has been crystallized and acquired. But in the process of acquisition it will function as an index facilitating the gradual crystallization of the novel meaning (concept) together with additional cues from the surrounding context and the recipient’s general background knowledge as illustrated in figure 2 (b) above. In the case of novel food names the contextual bias is provided, first and foremost, by other labelling elements on the same packaging, i.e. texts, figures, images, background colours, etc. (Smith 2009; Smith, Zlatev & Barratt forthcoming).

This adds new shades to current judgments on the misleading potential of novel food names which tend to circle around the name’s built-in semantic potential and its “objective” meaning. Consider Examples (6) – (9).

(6) **Halal ham**

Consumer insists that the product cannot be correctly named *halal skinke* ≈ ‘halal ham’ since it is made of turkey and not pork. While this may potentially compromise *skinke* ‘ham’, at the same time it justifies *halal*. Case No(s): 2005-01-274-00386 (id 319) CONSUMER; 2003-10-274-00462 (id 177) CONSUMER

(7) **Surimi shrimps**

New variant of minced fish meat product well-known to Danish consumers as *surimi* and originally designed to imitate crab meat. Complainant objects to the name *surimi-rejer* ≈ ‘surimi shrimps’ for a new, shrimp-shaped version because the product contains no shrimps. Case No(s): 2006-N4-274-00885 (id 506) CONSUMER

(8) **Pizzatop**

Product developed as functional (and sensory) equivalent to pizza cheese, but mainly containing other ingredients than cheese, marketed under the names *Pizzatop* and *Pizza Topping*. Main objection: Cheese is what you normally put “on top” of a pizza. Case No(s): 2006-N4-274-00998 (508) DAF; 2006-N4-274-00999 (509) DAF

(9) **Spread or butter?**

The name *Lurpak smørbar* ≈ 'Lurpak spreadable' on a blend product made of butter and vegetable oil claimed to be misleading for two interconnected reasons: 1. The brand name is also associated with "real" (high quality) butter, 2. the word *smørbar* ≈ 'spreadable' is etymologically related to *smør* ≈ 'butter' in Danish. Case No(s): 2006-06-274-00467 (458) CONSUMER

In (6) the alleged misleading potential comes down to the fact that *halal* and *ham* mutually exclude each other. However, it might alternatively be argued that precisely that conceptual clash might at best also support a non-misleading interpretation. This certainly requires a metaphorical extension of *ham* (just like of *yacht* in *land yacht* above), but if that is achieved, the constellation may indeed be a rather apt and compact way of conveying the following subtle message: This is as close as you get to something that looks, tastes, and feels like ham without disobeying **a religious proscription against eating pork**. A related case could be made for *surimi rejer* 'surimi shrimps' in (7) considering that many Danish consumers are familiar with "standard" *surimi* which, in turn, speaks against taking *shrimps* too literally. In cases like this, the misleading potential is certainly present, but the outcome is not determined by the meaning of the constituents alone. Ultimately, it will depend on how their relation is specifically interpreted in view of the consumer's general background knowledge, and, not least, by what has been done to support the intended interpretation through the context, *in casu*, the surrounding labelling – say, by claims like "great taste, no pork" or "now also in shrimp shape". In (8) and (9) there is even less in the built-in semantic potential of the names themselves that speaks against an interpretation which is fully consistent with facts, even without resorting to metaphorical extensions (though other interpretations are always possible, and a potentially misleading one is canonized as the "correct" one on etymological grounds in (9)). Since consumer knowledge can only be subject to general estimates, the key variable for ensuring the "right", i.e. the intended, interpretation (in view of such best estimates) therefore remains other words, texts, or images on the surrounding labelling. The name can *never* do the job alone, otherwise it would not be a name, but a definition.

Ongoing FairSpeak work includes the development of a schematized food labelling matrix in which 4 key biasing units (brand, verbal claim, illustration, colour(s)) can be varied systematically to test their joint potential of pushing the interpretation of a potentially ambiguous novel food name both in a misleading direction, and towards consumer understanding and acceptance (Smith, Zlatev & Barratt, forthcoming). The variables of interest here are not only semantic (communicative) potential, but also the capability of the various labelling elements to attract the consumer's visual attention. Whether this happens and the sequence in





*kammuslinger*, *hjertermuslinger*, etc. However, these do not all correspond to the Italian names mentioned in a one-to-one fashion. In short: Since the “bits and pieces” (word roots) available for coining new food names differ profoundly and systematically across languages, this is bound to affect the semantics of the food names already coined, and the ways in which novel names for novel products will be coined, interpreted, and gradually acquired by consumers. The manufacturer of surimi shrimps in example (7) might thus well invent yet another product variant and label it e.g. *surimi muslinger* for the Danish market. Yet naming it for the Italian market would take a different linguistic choice. While the present example may appear somewhat trivial (while being both realistic and illustrative), others are clearly not: With entirely new categories of foods being invented by European food manufacturers at a steady pace, including e.g. functional foods, special diet foods, etc., the challenge of naming them in a fair way while taking into account the different pre-established food naming patterns and language-system inherent typological preferences in different EU countries, and the food culture embedded in them, is all but trivial.

## 8. Concluding remarks

We have thus seen that the specifics (the substance) upon which a legal form<sup>8</sup> is imposed in real-life disputes on misleading food naming involve rather subtle linguistic and cognitive questions relating not only to the (conventionalized or alleged) meaning of the name as such, but also to the bias potentially exerted by other verbal and visual cues on the packaging when that meaning is to be established, and to the general knowledge and world views with which the individual consumer meets the task. Viewing these issues across national and linguistic/typological border adds yet other perspectives that constitute obvious subjects for future research.

There are thus good grounds for arguing that assessing the semantics of established and novel food names from a fairness perspective not only supports (societally highly essential) practical purposes, but can also serve as a catalyst for innovative research into the complexity of online word comprehension, lexicalization, and acquisition processes in general. That is: Into how the language-system (langue) is generated and steadily adjusted and developed in the actual language-behavior (parole) of particular individuals in particular real-life settings. *In casu*, when product developers and marketing specialists create and name new

---

<sup>8</sup> For further discussion on this application of the form/substance distinction (interpreted in Hjelmslev’s generalized sense, cfr. 1953 [1943]: 80-81) also to the law-language interface, see Smith (2007: 127-130).

products, and when the rest of us encounter and interpret the outcome during our daily shopping in our local supermarket.

## Bibliography

- Andrews, S. & Davis, C. J. 1999. Interactive activation accounts of morphological decomposition: Finding the trap in mousetrap? *Brain & Language* 68: 355-361.
- Barsalou, L.W. 1983. Ad hoc categories. *Memory & Cognition* 11(3): 211-227.
- Barsalou, L.W. 1987. The instability of graded structure: Implications for the nature of concepts. In U. Nesser (ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization*. Cambridge: Cambridge University Press, 101-139.
- Barsalou, L.W. 1999. Perceptual symbol systems. *Behavioral & Brain Sciences* 22(4): 577-609.
- Barsalou, L.W. 2005. Situated conceptualization. In H. Cohen, & C. Lefebvre (eds), *Handbook of categorization in cognitive science*. Amsterdam: Elsevier, 619-650.
- Benches, R. 2006. *Creative compounding in English*. Amsterdam: John Benjamins.
- Evans, V. & Green, M. 2006. *Cognitive linguistics: An introduction*. Mahwah (NJ): Lawrence Erlbaum.
- Fauconnier, G. & Turner, M. 2002. *The way we think: Conceptual blending and the mind's hidden complexities*. New York: Basic Books.
- Ferris, D.C. 1983. *Understanding semantics*. Exeter: University of Exeter.
- Gagné, C.L. 2001. Relation and lexical priming during the interpretation of noun-noun combinations. *Journal of Experimental Psychology: Learning, Memory and Cognition* 27: 236-54.
- Gagné, C.L., Spalding, T.L. & Gorrie, M.C. 2005. Sentential context and the interpretation of familiar open-compounds and novel modifier-noun phrases. *Language and Speech* 48(2): 203-221.
- Gill, T. & Dubé, L. 2007. What is a leather iron or a bird phone? Using conceptual combinations to generate and understand new product concepts. *Journal of Consumer Psychology* 17(3): 202-217.
- Gumperz, J.J. & Levinson, S.C. (eds) 1996. *Rethinking linguistic relativity*. Cambridge: Cambridge University Press.
- Hjelmslev, Louis 1953 [1943]. *Prolegomena to a Theory of Language*. Translated by Francis J. Whitfield. Bloomington (IN): Indiana University Press.
- Korzen, I. 2000. Pragmatica testuale e sintassi nominale. Gerarchie pragmatiche, determinazione nominale e relazioni anaforiche. In I. Korzen & C. Marellò (eds), *Argomenti per una linguistica della traduzione. Notes pour une linguistique de la traduction. On linguistic aspects of translation. Gli argomenti umani 4*. Alessandria: Edizioni dell'Orso, 81-109.
- Korzen, I. 2004. Dalla microstruttura alla macrostruttura. In P. D'Achille (ed.), *Generi, architetture e forme testuali. Atti del VII Convegno SILFI*. Firenze: Franco Cesati, 363-376.
- Krott, A. & Nicoladis, E. 2005. Large constituent families help children parse compounds. *Journal of Child Language* 32: 139-158.

- Libben, G. 2006. Why study compound processing? An overview of the issues. In G. Libben & G. Jarema (eds), *The representation and processing of compound words*. Oxford: Oxford University Press, 1-22.
- Libben, G. & Jarema, G. (eds) 2006. *The representation and processing of compound words*. Oxford: Oxford University Press.
- MacMaoláin, C. 2007. *EU food law. Protecting consumers and health in a common market*. Oxford: Hart Publishing.
- Manelis, L. & Tharp, D.A. 1977. The processing of affixed words. *Memory and Cognition* 5: 690-695.
- Moskowitz, H.R., Reisner, M., Itty, B., Katz, R., & Krieger, B. 2006. Steps towards a consumer-driven 'concept innovation machine' for food and drink. *Food Quality and Preference* 17(7-8): 536-551.
- Murphy, G.L. 1990. Noun phrase interpretation and conceptual combination. *Journal of Memory and Language* 29: 259-88.
- Møgelvang-Hansen, P. (in press). Misleading Presentation of Food. Methods of Legal Regulation and Real-Life Case Scenarios. In H.-W. Micklitz, V. Smith & M. Ohm Søndergaard (eds), *Supporting fairness in product-to-consumer communication - challenges in consumer protection. EUI Working Papers LAW*. Firenze: European University Institute.
- Ohm Søndergaard, M. & Selsøe Sørensen, H. 2008. Frugtsaft: Beskyttede varebetegnelser – et instrument til at undgå vildledning?. *Erhvervsjuridisk Tidsskrift* 4: 298-305.
- Piaget, J. 1926. *The language and thought of the child*. New York: Harcourt, Brace & World.
- Putnam, H. 1975. The meaning of 'meaning'. In *Mind, language, and reality. Philosophical Papers. Volume 2*. Cambridge: University Press, 215-271.
- Rosch, E. 1975. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* 104(3): 192-233.
- Sager, J.C. 1997. Term formation. In S.E. Wrigth & G. Budin (eds), *Handbook of terminology management. Volume 1. Basic aspects of terminology management*. Amsterdam: Benjamins, 25-41.
- Sandra, D. 1990. On the representation and processing of compound words: Automatic access does not occur. *Quarterly Journal of Experimental Psychology* 42A: 529-567.
- Schreuder, R. & Baayen, R.H. 1995. Modelling morphological processing. In L. Feldman (ed.), *Morphological aspects of language processing*. Hillsdale, NJ: Erlbaum Press, 131-154.
- Smith, E.E., Shoben, E.J. & Rips, L.J. 1973. Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review* 81(3): 214-241.
- Smith, V. 1999/2000. The literal meaning of lexical items. *Copenhagen Working Papers in LSP* 7.
- Smith, V. 2000. On the contrastive study of lexicalization patterns for translation purposes: some reflections on the levels of analysis. In I. Korzen & C. Marelllo (eds), *Argomenti per una linguistica della traduzione. Notes pour une linguistique de la traduction. On linguistic aspects of translation. Gli argomenti umani* 4. Alessandria: Edizioni dell'Orso, 19-42.

- Smith, V. 2001. Lexical non-arbitrariness and LSP translation: A causal model of explanation. *Fachsprache – International Journal of LSP* 1-2: 24-40.
- Smith, V. 2007. Linguistic diversity versus legal unity in Europe: Getting beyond the chicken-and-egg discussions. *The Journal of Comparative Law* 2(1): 120-136.
- Smith, V. 2009. Freezing the waves: From ad-hoc categorization to lexicalization on food naming. Power Point presentation at the 2<sup>nd</sup> Conference of the Swedish Association for Language and Cognition, Stockholm University, 10-12 June 2009. [http://www.fair-speak.org/PPTS/SALC\\_2009\\_PowerPoint\\_Smith\\_WEB-filer/frame.htm](http://www.fair-speak.org/PPTS/SALC_2009_PowerPoint_Smith_WEB-filer/frame.htm).
- Smith, V. (in press). What's in a (food) name? From consumer protection to cognitive science – and back. In H.-W. Micklitz, V. Smith & M. Ohm Søndergaard (eds), *Supporting fairness in product-to-consumer communication – challenges in consumer protection. EUI Working Papers LAW*. Firenze: European University Institute.
- Smith, V., Clement, J., Møgelvang-Hansen, P. & Selsøe Sørensen, H. (in review). Assessing in-store food-to-consumer communication from a fairness perspective: An integrated approach.
- Smith, V., Møgelvang-Hansen, P. & Hyldig, G. (in press). Spin versus fair speak in food labelling: A matter of taste? *Food Quality and Preference*.
- Smith, V., Søndergaard, M.O., Clement, J., Møgelvang-Hansen, P., Sørensen, H.S., & Gabrielsen, G. 2009. *Fair Speak: Scenarier for vildledning på det danske fødevaremarked*. Copenhagen: Ex Tuto Publishing.
- Smith, V., Zlatev, J. & Barratt, D. (forthcoming). Unpacking noun-noun compounds: Interpreting novel and conventional food names on plain paper and on simulated labels.
- Stekauer, P. 2006. On the meaning predictability of novel context-free converted naming units. *Linguistics* 44 (3): 489-539.
- Talmy, L. 2000. *Toward a cognitive semantics. Volume 1: Concept structuring systems*. Cambridge (MA): The MIT Press.
- Taylor, J.R. 1989. *Linguistic categorization: Prototypes in linguistic theory*. Oxford: Clarendon Press.
- Vygotsky, L.S. 1962 [1934]. *Thought and Language*. Cambridge (MA): MIT Press.
- Wierzbicka, A. 1985. *Lexicography and conceptual analysis*. Ann Arbor: Karoma.
- Wüster, E. 1959/60. Das Wort der Welt, Schaubildlich und terminologisch dargestellt. *Sprachforum* 3/4: 183-203.
- Zlatev, J., Smith, V., Weijer, J. van de, Skydsgaard, C. (in press). Noun-noun compounds for fictive food products: Experimenting in the borderzone of semantics and pragmatics. *Journal of Pragmatics*.



STRUMENTI  
PER LA DIDATTICA E LA RICERCA

1. Brunetto Chiarelli, Renzo Bigazzi, Luca Sineo (a cura di), *Alia: Antropologia di una comunità dell'entroterra siciliano*
2. Vincenzo Cavaliere, Dario Rosini, *Da amministratore a manager. Il dirigente pubblico nella gestione del personale: esperienze a confronto*
3. Carlo Biagini, *Information technology ed automazione del progetto*
4. Cosimo Chiarelli, Walter Pasini (a cura di), Paolo Mantegazza. *Medico, antropologo, viaggiatore*
5. Luca Solari, *Topics in Fluvial and Lagoon Morphodynamics*
6. Salvatore Cesario, Chiara Fredianelli, Alessandro Remorini, *Un pacchetto evidence based di tecniche cognitivo-comportamentali sui generis*
7. Marco Masseti, *Uomini e (non solo) topi. Gli animali domestici e la fauna antropocora*
8. Simone Margherini (a cura di), *BIL Bibliografia Informatizzata Leopardiana 1815-1999: manuale d'uso ver. 1.0*
9. Paolo Puma, *Disegno dell'architettura. Appunti per la didattica*
10. Antonio Calvani (a cura di), *Innovazione tecnologica e cambiamento dell'università. Verso l'università virtuale*
11. Leonardo Casini, Enrico Marone, Silvio Menghini, *La riforma della Politica Agricola Comunitaria e la filiera olivicolo-olearia italiana*
12. Salvatore Cesario, *L'ultima a dover morire è la speranza. Tentativi di narrativa autobiografica e di "autobiografia assistita"*
13. Alessandro Bertirotti, *L'uomo, il suono e la musica*
14. Maria Antonietta Rovida, *Palazzi senesi tra '600 e '700. Modelli abitativi e architettura tra tradizione e innovazione*
15. Simone Guercini, Roberto Piovon, *Schemi di negoziato e tecniche di comunicazione per il tessile e abbigliamento*
16. Antonio Calvani, *Technological innovation and change in the university. Moving towards the Virtual University*
17. Paolo Emilio Pecorella, *Tell Barri/Kahat: la campagna del 2000. Relazione preliminare*
18. Marta Chevanne, *Appunti di Patologia Generale. Corso di laurea in Tecniche di Radiologia Medica per Immagini e Radioterapia*
19. Paolo Ventura, *Città e stazione ferroviaria*
20. Nicola Spinosi, *Critica sociale e individuazione*
21. Roberto Ventura (a cura di), *Dalla misurazione dei servizi alla customer satisfaction*
22. Dimitra Babalis (a cura di), *Ecological Design for an Effective Urban Regeneration*
23. Massimo Papini, Debora Tringali (a cura di), *Il pupazzo di garza. L'esperienza della malattia potenzialmente mortale nei bambini e negli adolescenti*
24. Manlio Marchetta, *La progettazione della città portuale. Sperimentazioni didattiche per una nuova Livorno*
25. Fabrizio F.V. Arrigoni, *Note su progetto e metropoli*
26. Leonardo Casini, Enrico Marone, Silvio Menghini, *OCM seminativi: tendenze evolutive e assetto territoriale*
27. Pecorella Paolo Emilio, Raffaella Pierobon Benoit, *Tell Barri/Kahat: la campagna del 2001. Relazione preliminare*
28. Nicola Spinosi, *Wir Kinder. La questione del potere nelle relazioni adulti/bambini*
29. Stefano Cordero di Montezemolo, *I profili finanziari delle società vinicole*
30. Luca Bagnoli, Maurizio Catalano, *Il bilancio sociale degli enti non profit: esperienze toscane*
31. Elena Rotelli, *Il capitolo della cattedrale di Firenze dalle origini al XV secolo*
32. Leonardo Trisciuzzi, Barbara Sandrucci, Tamara Zappaterra, *Il recupero del sé attraverso l'autobiografia*
33. Nicola Spinosi, *Invito alla psicologia sociale*
34. Raffaele Moschillo, *Laboratorio di disegno. Esercitazioni guidate al disegno di arredo*
35. Niccolò Bellanca, *Le emergenze umanitarie complesse. Un'introduzione*
36. Giovanni Allegretti, *Porto Alegre una biografia territoriale. Ricercando la qualità urbana a partire dal patrimonio sociale*
37. Riccardo Passeri, Leonardo Quagliotti, Christian Simoni, *Procedure concorsuali e governo dell'impresa artigiana in Toscana*
38. Nicola Spinosi, *Un soffitto viola. Psicoterapia, formazione, autobiografia*
39. Tommaso Urso, *Una biblioteca in divenire. La biblioteca della Facoltà di Lettere dalla penna all'elaboratore. Seconda edizione rivista e accresciuta*
40. Paolo Emilio Pecorella, Raffaella Pierobon Benoit, *Tell Barri/Kahat: la campagna del 2002. Relazione preliminare*
41. Antonio Pellicanò, *Da Galileo Galilei a Cosimo Noferi: verso una nuova scienza. Un inedito trattato galileiano di architettura nella Firenze del 1650*

42. Aldo Burresti (a cura di), *Il marketing della moda. Temi emergenti nel tessile-abbigliamento*
43. Curzio Cipriani, *Appunti di museologia naturalistica*
44. Fabrizio F.V. Arrigoni, *Incipit. Esercizi di composizione architettonica*
45. Roberta Gentile, Stefano Mancuso, Silvia Martelli, Simona Rizzitelli, *Il Giardino di Villa Corsini a Mezzomonte. Descrizione dello stato di fatto e proposta di restauro conservativo*
46. Arnaldo Nesti, Alba Scarpellini (a cura di), *Mondo democristiano, mondo cattolico nel secondo Novecento italiano*
47. Stefano Alessandri, *Sintesi e discussioni su temi di chimica generale*
48. Gianni Galeota (a cura di), *Traslocare, ri-aggregare, rifondare. Il caso della Biblioteca di Scienze Sociali dell'Università di Firenze*
49. Gianni Cavallina, *Nuove città antichi segni. Tre esperienze didattiche*
50. Bruno Zanoni, *Tecnologia alimentare 1. La classe delle operazioni unitarie di disidratazione per la conservazione dei prodotti alimentari*
51. Gianfranco Martiello, *La tutela penale del capitale sociale nelle società per azioni*
52. Salvatore Cingari (a cura di), *Cultura democratica e istituzioni rappresentative. Due esempi a confronto: Italia e Romania*
53. Laura Leonardi (a cura di), *Il distretto delle donne*
54. Cristina Delogu (a cura di), *Tecnologia per il web learning. Realtà e scenari*
55. Luca Bagnoli (a cura di), *La lettura dei bilanci delle Organizzazioni di Volontariato toscane nel biennio 2004-2005*
56. Lorenzo Grifone Baglioni (a cura di), *Una generazione che cambia. Civismo, solidarietà e nuove incertezze dei giovani della provincia di Firenze*
57. Monica Bolognesi, Laura Donati, Gabriella Granatiero, *Acque e territorio. Progetti e regole per la qualità dell'abitare*
58. Carlo Natali, Daniela Poli (a cura di), *Città e territori da vivere oggi e domani. Il contributo scientifico delle tesi di laurea*
59. Riccardo Passeri, *Valutazioni imprenditoriali per la successione nell'impresa familiare*
60. Brunetto Chiarelli, Alberto Simonetta, *Storia dei musei naturalistici fiorentini*
61. Gianfranco Bettin Lattes, Marco Bontempini (a cura di), *Generazione Erasmus? L'identità europea tra vissuto e istituzioni*
62. Paolo Emilio Pecorella, Raffaella Pierobon Benoit, *Tell Barri / Kahat. La campagna del 2003*
63. Fabrizio F.V. Arrigoni, *Il cervello delle passioni. Dieci tesi di Adolfo Natalini*
64. Saverio Pisaniello, *Esistenza minima. Stanze, spazi della mente, reliquiario*
65. Maria Antonietta Rovida (a cura di), *Fonti per la storia dell'architettura, della città, del territorio*
66. Ornella De Zordo, *Saggi di anglistica e americanistica. Temi e prospettive di ricerca*
67. Chiara Favilli, Maria Paola Monaco, *Materiali per lo studio del diritto antidiscriminatorio*
68. Paolo Emilio Pecorella, Raffaella Pierobon Benoit, *Tell Barri / Kahat. La campagna del 2004*
69. Emanuela Caldognetto Magno, Federica Cavicchio, *Aspetti emotivi e relazionali nell'e-learning*
70. Marco Masetti, *Uomini e (non solo) topi (2ª edizione)*
71. Giovanni Nerli, Marco Pierini, *Costruzione di macchine*
72. Lorenzo Viviani, *L'Europa dei partiti. Per una sociologia dei partiti politici nel processo di integrazione europea*
73. Teresa Crespellani, *Terremoto e ricerca. Un percorso scientifico condiviso per la caratterizzazione del comportamento sismico di alcuni depositi italiani*
74. Fabrizio F.V. Arrigoni, *Cava. Architettura in "ars marmoris"*
75. Ernesto Tavoletti, *Higher Education and Local Economic Development*
76. Carmelo Calabrò, *Liberalismo, democrazia, socialismo. L'itinerario di Carlo Rosselli (1917-1930)*
77. Luca Bagnoli, Massimo Cini (a cura di), *La cooperazione sociale nell'area metropolitana fiorentina. Una lettura dei bilanci d'esercizio delle cooperative sociali di Firenze, Pistoia e Prato nel quadriennio 2004-2007*
78. Lamberto Ippolito, *La villa del Novecento*
79. Cosimo Di Bari, *A passo di critica. Il modello di Media Education nell'opera di Umberto Eco*
80. Leonardo Chiesi (a cura di), *Identità sociale e territorio. Il Montalbano*
81. Piero Degl'Innocenti, *Cinquant'anni, cento chiese. L'edilizia di culto nelle diocesi di Firenze, Prato e Fiesole (1946-2000)*
82. Giancarlo Paba, Anna Lisa Pecoriello, Camilla Perrone, Francesca Rispoli, *Partecipazione in Toscana: interpretazioni e racconti*
83. Alberto Magnaghi, Sara Giacomozzi (a cura di), *Un fiume per il territorio. Indirizzi progettuali per il parco fluviale del Valdarno empoiese*



84. Dino Costantini (a cura di), *Multiculturalismo alla francese?*
85. Alessandro Viviani (a cura di), *Firms and System Competitiveness in Italy*
86. Paolo Fabiani, *The Philosophy of the Imagination in Vico and Malebranche*
87. Carmelo Calabrò, *Liberalismo, democrazia, socialismo. L'itinerario di Carlo Rosselli*
88. David Fanfani (a cura di), *Pianificare tra città e campagna. Scenari, attori e progetti di nuova ruralità per il territorio di Prato*
89. Massimo Papini (a cura di), *L'ultima cura. I vissuti degli operatori in due reparti di oncologia pediatrica*
90. Raffaella Cerica, *Cultura Organizzativa e Performance economico-finanziarie*
91. Alessandra Lorini, Duccio Basosi (a cura di), *Cuba in the World, the World in Cuba*
92. Marco Goldoni, *La dottrina costituzionale di Sieyès*
93. Francesca Di Donato, *La scienza e la rete. L'uso pubblico della ragione nell'età del Web*
94. Serena Vicari Haddock, Marianna D'Ovidio, *Brand-building: the creative city. A critical look at current concepts and practices*
95. Ornella De Zordo (a cura di), *Saggi di Anglistica e Americanistica. Ricerche in corso*
96. Massimo Moneglia, Alessandro Panunzi (edited by), *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*
97. Alessandro Panunzi, *La variazione semantica del verbo essere nell'Italiano parlato*
98. Matteo Gerlini, *Sansone e la Guerra fredda. La capacità nucleare israeliana fra le due superpotenze (1953-1963)*
99. Luca Raffini, *La democrazia in mutamento: dallo Stato-nazione all'Europa*
100. Gianfranco Bandini (a cura di), *noi-loro. Storia e attualità della relazione educativa fra adulti e bambini*
101. Anna Taglioli, *Il mondo degli altri. Territori e orizzonti sociologici del cosmopolitismo*
102. Gianni Angelucci, Luisa Vierucci (a cura di), *Il diritto internazionale umanitario e la guerra aerea. Scritti scelti*
103. Giulia Mascagni, *Salute e disuguaglianze in Europa*
104. Elisabetta Cioni e Alberto Marinelli (a cura di), *Le reti della comunicazione politica. Tra televisioni e social network*
105. Cosimo Chiarelli, Walter Pasini (a cura di), *Paolo Mantegazza e l'Evoluzionismo in Italia*
106. Andrea Simoncini (a cura di), *La semplificazione in Toscana. La legge n. 40 del 2009*
107. Claudio Borri, Claudio Mannini (edited by), *Aeroelastic phenomena and pedestrian-structure dynamic interaction on non-conventional bridges and footbridges*
108. Emiliano Scampoli, *Firenze, archeologia di una città (secoli I a.C. - XIII d.C.)*
109. Emanuela Cresti, Iørn Korzen (edited by), *Language, Cognition and Identity. Extensions of the endocentric/exocentric language typology*

