

Parole in Compagnia  
Presentazione del Modulo di  
Terminologia e Linguaggi Settoriali  
su Collocazioni e Altre Parole Associate

A.A. 2005/2006

### Informazioni generali

**Docente** Marco Baroni

**Email** baroni@sslmit.unibo.it

**Telefono** 0453/374744

**Homepage del corso** <http://sslmit.unibo.it/~baroni/collocazioni>

**Orario lezioni interpreti** Martedì 9:00-10:30 in aula C1; venerdì 12:00-13:30  
in aula D2 (tot. 24 ore)

**Orario lezioni traduttori** Venerdì 9:00-10:30 in aula D2 (tot. 14 ore)

**Ricevimento** Venerdì 10:30-12:00 in Aula Rosiello, e su appuntamento

### Collocazioni & Co.

- You should tell a word by the company it keeps (Firth)
- Competenza linguistica non riguarda solo conoscenza di parole e regole combinatorie generali (sintassi, morfologia), ma anche memorizzazione delle preferenze che regolano come coppie/gruppi di parole specifici si combinano
- Esempi:
  - tirare le cuoia; menare il can per l'aia
  - caffè forte (\*caffè potente); computer potente (\*computer forte)

- prendere una decisione vs. make a decision; avere senso vs. make sense
- portare pazienza; portare fortuna
- fare una telefonata; fare il bagno
- latte e miele; cane e gatto
- causare danni, gravi difficoltà, ??gioia
- vaso sanguigno; sospensione orale
- pesce spada; anima gemella
- nella misura in cui; onorevoli colleghi
- discussione di, su, riguardo a
- finire un libro; finire il dolce
- leggere un libro, Chomsky; spolverare un libro; invitare Chomsky; divorare Chomsky
- parcheggiare la macchina; cantare una canzone

## Motivazioni teoriche

- Tradizione linguistica strutturalista/generativa pone (giustamente!) enfasi su aspetti combinatori del linguaggio
- Regole sintattiche formano scheletro di caselle vuote, che possono venire riempite liberamente da elementi della categoria grammaticale giusta (e.g.,  $S \rightarrow NP VP$ ;  $NP \rightarrow Det Adj Noun$ )
- Lessico contiene solo parole o morfemi, elementi minimi non ulteriormente analizzabili che vengono inseriti negli scheletri vuoti creati da regole sintattiche
- Più di recente, molte teorie (construction grammars, simpler syntax, HP-SG. . .) e studiosi (Fillmore, Jackendoff, Croft) hanno richiamato l'attenzione su vasta area "intermedia" tra lessico e sintassi
- Idiomi e collocazioni come "frammenti" sintagmatici immagazzinati nel lessico
- Continuum, da regole astratte a singoli morfemi, di frammenti di varia complessità e astrattezza, tutti immagazzinati nel lessico
- Necessità di studi che sistematizzino il panorama di collocazioni, idiomi, ecc.
- Legami tra parole hanno ruolo fondamentale anche nello studio di polisemia, costruzione del significato nel contesto, disambiguazione di sensi ecc. (Pustejovsky)

## Motivazioni pratiche

- A livello avanzato di conoscenza di lingua straniera, collocazioni fanno la differenza (“inglese buono ma poco idiomatico”)
- Anche in lingua madre, soprattutto in condizioni anomale per sistema linguistico (e.g., influenza di altra lingua durante una traduzione), talvolta “manca la collocazione”, ed esistono pericolosi falsi amici collocazionali (“fare senso”)
- “Traduzionese” caratterizzato anche da profilo collocazionale significativamente diverso da quello di testi originali comparabili (Baroni & Bernardini, *Proceedings of Corpus Linguistics 2003*)
- Con alcune fortunate eccezioni (BBI) mancano risorse di riferimento per le collocazioni
- Studio sistematico di collocazioni utile anche per
  - lessicografia tradizionale (fraseologia nelle voci di dizionario)
  - terminologia:
    - \* gran parte dei termini in term-bases sono termini composti, ma spesso manca approccio sistematico a distinzione tra tipi di termini semplici e composti
    - \* fraseologia/collocazioni fanno parte di conoscenza di lingua settoriale tanto quanto termini – terminologia tradizionale spesso ignora verbi probabilmente proprio in quanto elementi inerentemente relazionali

## Le collocazioni alla SSLMIT

- Cristofori/Valente: tesi di laurea specialistiche su prototipo di un dizionario combinatorio online per l'italiano
- Progetto “fraseologia per gli interpreti” di Amantia, Bertaccini, Mack, Prandi e Zanchetta
- Silvia Bernardini: tesi di dottorato sulle collocazioni nei testi tradotti
- Sara Piccioni: collocazioni nei testi letterari (in particolare, García Lorca)

## Il corso

### Programma (ovviamente provvisorio)

1. Questioni metodologiche:
  - I corpora
  - Estrazione di collocazioni da corpora
2. Aspetti teorici:
  - Tipologia e proprietà caratterizzanti di idiomi, collocazioni, termini, ecc.
3. Analisi dei dati (lavoro di gruppo):
  - Studio, classificazione e sistematizzazione di campioni di collocazioni estratti da corpora
  - Patterns: Verbo Nome; Aggettivo Nome (Nome Aggettivo)
  - Focus su italiano e inglese
  - Corpora:
    - Generali
    - Linguaggi speciali
    - Letteratura
    - Discorsi
4. Fase applicativa:
  - Progettazione di un prototipo di database flessibile delle collocazioni in italiano (da rendere liberamente disponibile in rete)

### Lecture

- Obbligatorie (disponibili in via Oberdan):
  - E. Ježek. 2005. Lessico. Bologna: il Mulino. Sezione 3.4 del capitolo 3 e capitolo 6
  - C. Manning & H. Schütze. 1999. Foundations of statistical natural language processing. Boston: MIT Press. Sezioni 5, 5.1, 5.2, 5.5 del capitolo 5
- Approfondimenti:
  - S. Bartsch. 2004. Structural and functional properties of collocations in English.
  - A. Lenci, S. Montemagni & V. Pirrelli. 2005. Testo e computer. Roma: Carocci. In particolare, i capitoli 1, 2, sezioni 4 e 4.1 del capitolo 4, sezioni 7, 7.1, 7.2 e 7.4 (sulle collocazioni) del capitolo 7

- Il libro della Ježek nella sua interezza, come introduzione alle questioni legate al lessico e alla semantica lessicale in genere
- M. Benson, E. Benson & R. Ilson. 1986. *The BBI combinatory dictionary of English*. Amsterdam: Benjamins. Sezioni introduttive (in particolare, le parti sulle collocazioni lessicali).
- S. Evert. 2005. *The statistics of word cooccurrences: Word pairs and collocations*. Stuttgart: IMS.
- A. Cowie (a cura di). 1998. *Phraseology: Theory, analysis and applications*. Oxford: Clarendon. In particolare, gli articoli di Mel'čuk, Moon, Howarth, Fontenelle e Cowie.
- G. Nunberg, I. Sag & T. Wasow. 1994. *Idioms*. *Language* 70: 491-538.
- P. Howarth. 1996. *Phraseology in English academic writing*. Tübingen: Niemeyer. Capitolo 2
- N. Nesselhauf. 2005. *Collocations in a learner corpus*. Amsterdam: Benjamins.
- L. Grant & L. Bauer. 2004. *Criteria for re-defining idioms: Are we barking up the wrong tree?* *Applied Linguistics* 25(1): 38-61
- L. Grant. 2005. *Frequency of 'core idioms' in the British National Corpus (BNC)*. *International Journal of Corpus Linguistics*. 10: 429-451.
- M. Wynne (a cura di). 2005. *Developing linguistic corpora: a guide to good practice*. Oxford: Oxbow. In rete all'indirizzo <http://ahds.ac.uk/linguistic-corpora/>. In particolare, i due capitoli di Sinclair, e i capitoli di Leech e McEnery/Xiao.
- A. Cruse. 1985. *Lexical semantics*. Cambridge: CUP. In particolare la sezione 4.12 e le pp. 278-285.
- A. Bisetto. 2003. *Composizione con elementi italiani*. In M. Grossmann & F. Rainer (a cura di), *La formazione delle parole in italiano*. Tübingen: Niemeyer, pp. 33-69 (per la questione della delimitazione di composti e collocazioni/idiomi)
- Altri materiali verranno eventualmente segnalati a lezione

## Valutazione

1. Breve relazione finale sull'analisi condotta
2. Discussione delle letture obbligatorie, della relazione, e analisi orale di una lista di collocazioni