

Verbal lexicon emergency in Brazilian Portuguese

Mirian Rose Brum-de-Paula, Universidade Federal de Pelota
Giovana Ferreira-Gonçalves, Universidade Federal de Pelota

Differently from other elements of language, subjects are always able to acquire new lexical items. Such complex lexical development process occurs in an irregular way, actually, non linear development is an expected pattern of emergent systems. The development of verbal lexicon is cognitive and it is linguistically more complex than noun lexicon. Following an emergentist perspective of language acquisition process, according to which grammar patterns emerge from lexicon; this paper aims to reflect upon the way through which verbal lexicon occurs. Thus, we compared the lexicon produced by 8 monolingual children under the ages of 1 to 4 years-old who were acquiring Brazilian Portuguese as their first language. We established relations between frequency of lexical items and verbal morphology emergency. Data collections were longitudinal, consisting of care taker-child interactions and each of them lasted approximately 30 minutes. Data were transcribed and, afterwards, submitted to a percent survey on frequency performed by WordSmith Tools software. Results indicate some trends: verbal forms associated to present tense, infinitive and imperative modes; nuclear verbs, that is, those that allow lexical instantiation of great semantic domains by a frequent and underspecified unit; and lexical forms that are directly related to input frequency. In the present work, we defend an integrative conception of language acquisition, that is, the development and acquisition of language result from the product of the different components of language capacity, such as phonology and lexicon; morphology and lexicon, especially in this research.

Conventionalised verbal constructions in winespeak

Valeria Caruso, Università di Napoli "L'Orientale"

Anna Riccio, Università di Napoli "L'Orientale"

In the present study, we investigate the conventionalized verbal constructions of winespeak in a comparable multilingual corpus (English, French, Italian and German) of testing notes and wine reviews. We present quantitative estimations of the sensory domains (taste, smell, touch, hearing, sight) lexicalized by these units, as well as a more in-depth semantic-syntactic analysis of these constructions. For example, in the following sentences (1a-d) the verbal constructions convey meanings related to the physical perception of the sense organ, as well as to the internal sensations experienced by the perceiver:

(1)

- a. [...] a powerful wine that is nonetheless soft and ample and fills the mouth with ripe rather than aggressive tannin
- b. *Le vin remplit la bouche et les saveurs s'éclatent avec toujours ces tanins de velours [...]*
- c. *Era un vino che riempiva il palato, accarezzandolo anche pesantemente.*
- d. *Der Wein füllt den Mund mit einer enormen Kirschfrucht [...]*

The corpus analysis will be carried out according to the principles of the Frame Semantics (Fillmore 1977a, 1982, 1985; Fillmore and Atkins 1992) and following the FrameNet annotation and description methodology (Ruppenhofer *et al.* 2010), since semantic frames are useful "interlingual representations" for cross-linguistic comparisons (Boas 2005). Moreover, the Frames offer valuable descriptions for verbal metaphors (see in particular Dobrovolskij & Piirainen 2005), which can be depicted in terms of source and target concepts, and as such, using source and target frames.

Specific attention will be also paid to the semantic description of the perception modality and intensity, and how these are portrayed within the Frames of the FrameNet project. The overall results of the investigation will focus on the way intensity is lexicalized within the collected constructions and the corresponding descriptions offered by the FrameNet.

References

- Boas H. (2005). Semantic Frames as Interlingual Representations for Multilingual Lexical Databases. *International Journal of Lexicography* 18(4): 39-65.
- Dobrovolskij D & Piirainen E. (2005). *Figurative Language: Cross-cultural and cross-linguistic Perspectives*. Elsevier, Oxford.
- Fillmore C. (1977a). Scenes-and-Frames Semantics, Linguistic Structures Processing. In A. Zampolli (ed.), *Fundamental Studies in Computer Science* (59): 55-88.
- Fillmore C. (1982). Frame Semantics. In *Linguistics in the Morning Calm*, Seoul: Hanshin Publishing Co: 111-137.
- Fillmore C. (1985). Frames and the Semantics of Understanding. *Quaderni di Semantica* 6(2): 222-254.
- Fillmore C. & Atkins S. (1992). Towards a Frame-based Lexicon: The semantics of RISK and its Neighbors. In A. Lehrer & E. Kittay (eds.), *Frames, Fields, and Contrast: New Essays in Semantics and Lexical Organization*. Hillsdale: Lawrence Erlbaum Associates: 75-102.
- Ruppenhofer J., Ellsworth P., Petruck M., Johnson C. & Scheffczyk J. (2010). *FrameNet II: Extended Theory and Practice*. ICSI Technical Report.

Collocations in corpora and in the (linguists') mind: Combining corpus and experimental evidence

Adriano Ferraresi, Università di Bologna

Research on collocation has occupied centre stage in corpus linguistics since its early years. In line with the focus on the study of language performance “as a product” (Leech, 1992:108), attention has been placed on studying collocation as a textual phenomenon, i.e. as a tendency of words to frequently co-occur in corpora. It is only more recently that the “process”, or competence-related counterpart of the phenomenon has begun to be explored within the discipline, with a focus on understanding the relationship between corpus data and the representation of word combinations in the speakers' mental lexicon (Gilquin & Gries, 2009).

The proposed contribution aims to shed further light on this relationship by investigating the extent to which different lexical Association Measures (AMs) produce results matching human intuition on the salience of word combinations. Unlike other studies pursuing this line of research, the present work does not set a priori criteria to establish what a “salient collocation” is (as in e.g. Evert, 2005: chapter 5), and draws on judgments produced by expert linguist informants, rather than more linguistically naïve subjects (as in e.g. Siyanova & Schmitt, 2008). This is done with a view to answering two inter-related questions:

- a) What AM, if any, better matches expert judgements on the salience of (corpus-derived) collocations?
- b) Is it possible to identify shared criteria underlying such judgments?

Focusing on adjective-noun pairs, collocation candidates are extracted from a specialized corpus of institutional academic texts, a domain expected to be familiar to the study subjects. The pairs are scored using three among the most widely used AMs, i.e. bare frequency of co-occurrence, Log-Likelihood (LL) and Mutual Information (MI), and a relatively new one, i.e. Lexical Gravity (LEXG). A sample of the “top” word pairs according to each AM in three frequency ranges (high, medium and low), resulting in a dataset of 99 items, is then submitted in the form of an acceptability judgement questionnaire to participants in a corpus linguistics conference, who are asked to evaluate the collocation candidates in terms of how salient or lexically associated they perceive them to be, and to motivate their choice.

The poster presents quantitative results indicating that none of the AMs clearly outperforms the others in selecting salient word pairs. However, a clear distinction emerges whereby bare frequency, LL and LEXG correlate more strongly with the expert judgments, but MI is better at selecting the most salient pairs. The outcomes of a more qualitative analysis, triangulating results with lexicographic evidence are also presented. They suggest that, despite the limited consensus, the judgments seem to be guided by a shared set of lexico-syntactic and semantic criteria, classifying word combinations along a scale of cohesiveness, from compositional expressions to compounds. The main aim is to stimulate discussion on the implications of these findings and how they relate to those obtained in previous studies: taken together, these confirm that AMs *do* extract collocations from corpora matching expert intuition, but also indicate that more work is needed to better align performance- and competence-based perspectives.

References

- Gilquin G. & Gries S. T. (2009). Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory*, 5(1): 1-26.
- Leech G. (1992). Corpora and theories of linguistic performance. In J. Svartvik (ed.) *Directions in corpus linguistics*. Berlin: Mouton de Gruyter: 105-126.
- Siyanova A. & Schmitt N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *The Canadian Modern Language Review*, 64(3):429-458.
- Evert S. (2005). *The statistics of word cooccurrences: Word pairs and collocations*. PhD thesis, Universität Stuttgart, Stuttgart.

Il linguaggio politico: lessico e argomentazione

Francesca Ferrucci, Università degli Studi della Toscana

L'obiettivo di questa ricerca sul linguaggio politico è duplice: in primo luogo, sul piano teorico, verificare la possibilità di formulare una definizione unitaria dei molti fatti diversi raccolti sotto questa denominazione (discorsi orali, scritti, video trasmessi, prodotti da esponenti istituzionali, commentatori, etc.); in secondo luogo, sul piano applicativo, spiegare il problema di come esso si collochi rispetto alle altre modalità della variazione all'interno delle nozioni di 'linguaggio specialistico', 'tecnico-scientifico', 'settoriale'. In questa prospettiva, partendo dalle liste lessicografiche esistenti in ambito politico, elaborate a scopi e con approcci diversi, si è costruito un lemmario organico che raccoglie: a) termini esclusivamente politici (es. *premierato*, *referendum*); b) accezioni tecnico-politiche di parole diffuse anche nel vocabolario comune (es. *accordo*, *crisi*); c) locuzioni polirematiche tecniche politiche (es. *campagna elettorale*, *opinione pubblica*). Si è così ottenuto un lemmario tecnico-politico di 6.261 entrate, collegato da vincoli referenziali all'ambito di applicazione specifico costituito dalle attività politiche, amministrative e istituzionali.

Secondo il metodo della linguistica dei corpora e dell'analisi automatica dei testi, si è utilizzato il lemmario per interrogare corpora con caratteri diversi sia di tipo pragmatico (posizione sociale di mittente e destinatario, simmetria/asimmetria della comunicazione), sia sugli assi diamesico e diafasico. Come primo dato quantitativo è stata calcolata l'incidenza percentuale del lemmario: i risultati mostrano che, a parità di dimensioni dei campioni esaminati, l'uso del lessico politico è molto maggiore (doppio, o perfino triplo) in alcuni corpora rispetto ad altri, che trattano argomenti non esclusivi e sembrano ispirati da obiettivi comunicativi non strettamente legati alla riflessione sulla cosa pubblica. Ciò permette di associare una misura della frequenza lessicale a un maggiore o minore grado di riconoscibilità del genere specialistico politico. In secondo luogo, si sono presi in esame i modi con cui è utilizzato il lemmario: a differenza che nelle terminologie scientifiche, esso non presenta sempre univocità semantica nell'uso vivo dei parlanti, mostrando invece delle oscillazioni dei sensi concretamente realizzati rispetto alla definizione lessicografica. Nel 35,21% delle concordanze si è riscontrata questa variabilità, che risente del punto di vista della parte politica a cui il mittente aderisce, e che intende sostenere. L'immagine del genere discorsivo politico che ne risulta è ben distante dalle testualità descrittive-informative dei linguaggi scientifici e, viceversa, fortemente condizionata dalla finalità persuasiva – dato il carattere specifico del discorso politico (manifestazione di ipotesi di parte su argomenti di interesse collettivo) –, che si manifesta in un uso peculiare della terminologia politica. Tale uso consiste nella frequente ridefinizione del valore semantico fissato nelle fonti, che il parlante politico attua ricorrendo a procedure e dispositivi pragmatici (atti espositivi e commenti epilinguistici, implicature, attivazioni di presupposizioni).

Il poster presenta il profilo unitario del linguaggio politico risultante dalle analisi empiriche, enfatizzando la peculiarità del metodo che, a differenza di quanto avviene in molta letteratura specialistica, ha lo scopo di indagare non i caratteri specifici del linguaggio di un esponente, partito o periodo politico, ma di cogliere quanto caratterizza nel suo complesso quest'ambito della comunicazione pubblica: si parte dall'eterogeneità degli usi per estrarre alcuni tratti comuni, assegnando quindi agli usi stessi uno status di fondamento della teoria. L'esposizione è accompagnata da grafici su: 1) composizione statistica del lemmario politico secondo diversi parametri; 2) incidenza quantitativa del lemmario su diversi corpora; 3) esemplificazioni delle oscillazioni di senso con cui il lemmario è utilizzato, per alcune parole-chiave. Per quest'ultime si analizzano le principali modalità di negoziazione, nell'ambito di strategie retoriche e argomentative più complessive, fornendo esempi di concordanze.

La grammaticalizzazione dei costrutti (bi)nominali SN1 di SN2 nei dialetti italo-romanzi

Lorena Feudo, Università degli Studi di Firenze | Ruprecht-Karls-Universität Heidelberg

Negli ultimi decenni gli studi sulla teoria della grammaticalizzazione sono stati sempre più numerosi e hanno interessato diversi aspetti della linguistica: dalla linguistica storica, tipologica fino alla linguistica cognitiva e alla *Construction Grammar*. Ciononostante, poche, o comunque frammentarie, sono state le indagini sui fenomeni implicati nel processo di grammaticalizzazione soprattutto in ambito romanzo e dialettologico.

Il lavoro si propone di analizzare due casi: i quantificatori nei dialetti mediani e meridionali, e in particolare l'attenzione è rivolta alle costruzioni nominali SN1 di SN2 del tipo “un boccone di pane”, e lo sviluppo di *cria* come morfema della negazione postverbale. Nell'italiano moderno e soprattutto nelle varietà settentrionali, è ormai accertato l'uso di *mica* e di altre varianti (*brisa*) sia come quantificatori che come marche della negazione; quest'uso è emerso anche nelle varietà analizzate, dove nomi che inizialmente avevano un pieno significato lessicale – *cica/zica*, *boccone*, *tozzo*, *cria* – hanno subito un indebolimento semantico (*semantic bleaching*) e delle riduzioni fonetiche fino a diventare quantificatori, avverbi e complementi. Attraverso processi di grammaticalizzazione e partendo da un meccanismo di metonimia “la parte per il tutto”, il percorso è di solito avvenuto attraverso cinque stadi, ripresi dagli studi di Traugott (2008) e Brems (2011) sulle costruzioni inglesi *a bit of* e simili: 1) uso partitivo limitato (*Na cica de pane*); 2) uso partitivo esteso (*na cica de vino*); 3) Quantificatore (*Ma pe na zica de soddisfazione*); 4) Avverbio/modificatore scalare (*il padre era neccone severo*); 5) Complemento (*Sénti zzica!*).

Come emerge dalla raccolta dei dati, non tutti i lessemi hanno però raggiunto lo stesso stadio di grammaticalizzazione e solo uno, *cria*, nel dialetto di Carovilli (Isernia), sembra presentare un ulteriore sviluppo: quello in *negative marker*, una particella negativa postverbale che accompagna la negazione principale, Neg1 + V + Neg2 (*a me ne m'è mancat cria*). Da questo punto di vista e come hanno evidenziato Bernini & Ramat (1992), la grammaticalizzazione di *cria* è simile a quella del francese *pas* o *mica* dell'italiano settentrionale e rientra nell'ormai noto ciclo di Jespersen, anche se come si vedrà, la varietà molisana presenta ancora entrambe le negazioni.

Basandoci sugli studi sulla grammaticalizzazione – Heine *et al.* (1991), Hopper & Traugott (1993) – ma facendo riferimento anche alle teorie che pongono l'attenzione sul concetto di “contesto” e “costruzione”, il lavoro vuole aggiungere nuovi dati allo studio di questi dialetti, fornendo soprattutto un *corpus* concreto al quale si potrà attingere in futuro. In particolare, il caso della particella negativa *cria* mostra che questo processo non è limitato solamente alle varietà settentrionali o al toscano, evidenziato da Garzonio (2008) con l'analisi di *punto*, ma è riscontrabile anche nelle zone mediane e meridionali.

Riferimenti

- Bernini G. & Ramat P. (1992) *La frase negativa nelle lingue d'Europa*. Bologna, Il Mulino.
- Brems L. (2011) *Layering of size and type noun constructions in English*, Berlin, Mouton de Gruyter.
- Garzonio J. (2008) Diacronia e sincronia della sintassi di *punto* in Fiorentino, in D. Pescarini & F. Cognola (a cura di): *La negazione: variazione dialettale ed evoluzione diacronia*. Quaderni di Lavoro ASIt 8: 21-42.
- Heine B., Claudi U. & Hünnemeyer F. (1991) *Grammaticalization. A conceptual framework*, Chicago and London, The University of Chicago Press.
- Hopper P. & Traugott Closs E. (1993) [2003] *Grammaticalization*, Cambridge, Cambridge University Press.
- Ramat P. (1987) *Linguistic Typology*, Berlin, Mouton de Gruyter: 165-187.
- Traugott E. (2003) “Constructions in grammaticalization”, in B.D. Joseph & R.D. Janda (eds.), *The handbook of Historical linguistics*, London, Blackwell: 624-647.
- Traugott E. C. (2008) The grammaticalization of NP of NP patterns, in A. Bergs & G. Diewald (eds) *Constructions and Language Change*, Berlin, Mouton de Gruyter: 23-45.

Interazione tra frequenza e schematismo nelle preposizioni complesse in italiano. Verso un modello esplicativo *usage-based*.

Vittorio Ganfi, *Università degli Studi Roma Tre*

Valentina Piunno, *Università degli Studi Roma Tre*

Nonostante alcuni recenti studi (*inter alia* Casadei, 2001; Veland, 2011), una caratterizzazione unitaria della classe delle preposizioni complesse italiane rimane piuttosto problematica. Le preposizioni complesse costituiscono infatti un insieme eterogeneo tanto per il tipo di struttura sintattica quanto per il grado di fissità e di lessicalizzazione (Giacalone Ramat, 1994; Voghera, 1994). Il caso delle preposizioni complesse merita inoltre particolare attenzione, data l'importanza quantitativa e l'alta frequenza d'uso (Voghera, 2004). Benché le preposizioni complesse mostrino una minore variabilità di struttura e siano meno numerose rispetto ad altri tipi di lessemi complessi (es. nominali o verbali), sono più frequenti in termini assoluti: avvicinandosi per funzione agli elementi grammaticali, presentano infatti una più ampia possibilità combinatoria. Le preposizioni complesse in italiano presentano le seguenti configurazioni:

- PREP + NOUN + PREP (es. *in grado di, da parte di*)
- PREP + DET + NOUN + PREP (es. *nei confronti di, sulla base di*)
- NOUN + PREP (es. *riguardo a, rispetto a*)
- ADV + PREP (es. *prima di, fino a, oltre a, insieme a*)
- PREP + DET + PREP + ADV + PREP (es. *al di fuori di, al di sotto di*)

In questo contributo metteremo in relazione la schematicità e la fissità delle strutture preposizionali con la loro frequenza di impiego. Si dimostrerà pertanto che esiste una relazione fra la cristallizzazione degli elementi costitutivi di una preposizione complessa e la frequenza con la quale la costruzione occorre.

L'indagine, condotta sulla base di dati estratti dal *corpus* di italiano scritto "La Repubblica" (Baroni, 2004), intende individuare i diversi gradi lessicalizzazione e schematicità delle sequenze preposizionali complesse. A tal fine ci avvarremo di specifici test sintattico-semantiche: a. inserzione del determinante (*a rischio di vs. al rischio di*); b. separabilità dei costituenti (es. *nel mezzo di vs. nel bel mezzo di*); c. omissione di uno dei costituenti (*in cambio di vs. in cambio*); d. sostituzione di un costituente con un possessivo (*al fianco di vs. al suo fianco*); e. variazione della preposizione (*a causa di vs. per causa di*); f. perdita del valore designativo dell'elemento nominale (es. *per mano di vs. con le sue mani*).

Si vedrà come le strutture più frequenti tendono ad assumere tratti di schematicità e ad essere impiegate nelle forme più vicine alla grammatica: la frequenza di impiego, fattore importante nei processi di grammaticalizzazione (Bybee, 1985), favorisce infatti il trinceramento di pattern sintattici (*entrenchment*; Langacker, 1987; Goldberg, 1995, 2006; Brooks & Tomasello, 1999).

Mostreremo quindi che l'utilizzo di un modello descrittivo *usage-based* (Bybee & Thompson, 1997; Bybee, 2011) permette di considerare in maniera unitaria le dinamiche coinvolte nella formazione delle strutture preposizionali complesse dell'italiano contemporaneo, integrando in un coerente modello di sviluppo i dati statistici –correlati all'uso– con il grado di schematicità delle costruzioni.

Riferimenti

Baroni M. *et al.* (2004) Introducing the La Repubblica corpus: a large, annotated, TEI (XML)-compliant corpus of newspaper Italian, in M.St. Lino *et al.* (eds.), *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, 2 (LREC 2004, Lisbon, may 26-28), Paris, ELRA - European Language Resources: 1771-1774.

Brooks P. & Tomasello M. (1999) How children constrain their argument structure constructions, *Language*, 75: 720–738.

Bybee J. (2006) From Usage to Grammar: The Mind's Response to Repetition, *Language*, 82(49): 711-733.

Bybee J (2011) Usage-based Theory and Grammaticalization, in H. Narro & B. Heine, Bernd (eds.), *The Oxford Handbook of Grammaticalization*, Oxford, Oxford University Press: 69-78.

Bybee J. & Thompson S. (1997) Three Frequency Effects in Syntax, in *Proceedings of the Twenty-Third Annual Meeting of the Berkeley*, Berkeley Linguistics Society, 23: 378-388.

- Casadei F. (2001) Le locuzioni preposizionali. Struttura lessicale e gradi di lessicalizzazione”, *Lingua e Stile*, 36(1): 21-36.
- Giacalone Ramat A. (1994) Fonti di grammaticalizzazione. Sulla ricategorizzazione di verbi e nomi come preposizioni, in P. Cipriano, P. Di Giovine & M. Mancini (eds.) *Miscellanea di studi linguistici in onore di Walter Belardi*, Roma, Il Calamo, vol. II, 877-896.
- Goldberg A. (1995) *Constructions. A construction Grammar Approach to Argument Structures*, Chicago, The University of Chicago Press.
- Goldberg A. (2006) *Constructions at work*, Oxford, Oxford University Press.
- Langacker R. W. (1987) *Foundations of cognitive grammar*, Stanford, Stanford University Press.
- Veland R. (2011) Le locuzioni preposizionali *a(d) opera di, per opera di, per mano di* in italiano contemporaneo: funzionamento sintattico e proprietà semantiche, *Zeitschrift für romanische Philologie*, 128(3): 537-552.
- Voghera M. (1994) Lessemi complessi: percorsi di lessicalizzazione a confronto, *Lingua e Stile*, 29(2): 185-214.
- Voghera M. (2004) Polirematiche, in M. Grossmann & F. Rainer (eds.), *La formazione delle parole in italiano*, Tübingen, Max Niemeyer Verlag, 56-69.

The notion of entrenchment: A psycholinguistic experiment on L1 and L2 processing of morphologically complex words

Hélène Giraud, CNRS & Université de Toulouse
Serena Dal Maso, Università di Verona

The research aims at investigating the degree of entrenchment of morphologically complex words in L1 and L2 users' competence and its effect on word processing. In usage-based models, an element is conceived as mentally entrenched when it is perceived as a unit which has cognitive precedence over its components (see Blumenthal-Dramé, 2012; but also Bybee, 2010 and Tomasello, 2003). As far as morphologically complex words are concerned, entrenchment is determined by the strength of the lexical representation of the complex words (as a whole) and of its morphological constituents. Lexical strength, in its turn, mainly depends on token frequency.

So far, psycholinguistic research has demonstrated that the probability for a morphologically complex word to be accessed by whole-word processing, or by the processing of its morphological components, is determined by the balance between the frequency of the complex word (as a whole) and the frequency of the morphemes present in the stimulus (Beauvillain, 1996; Burani & Caramazza, 1987; Colé *et al.*, 1989; Schreuder *et al.*, 2002).

For Italian, Burani & Thornton (2003) investigated the interplay between the frequency of the root, the frequency of the suffix and the whole-word frequency in processing derived words. More precisely, they considered low frequency suffixed words which differed with respect to the frequency of their morphemic constituents.

In the present study we will verify whether the different degree of entrenchment is determinant also in the processing of L2 morphologically complex words. This is particularly interesting, as some authors argue for the existence of substantial differences between native and non-native processing of morphologically complex words (Silva & Clahsen, 2008), while others (Dal Maso & Giraud, *in press*) found morphological priming effects also in L2 processing, at least for words having a high surface frequency and very productive suffixes. In order to further investigate the role of morphology in the L2 processing and more specifically to verify the whole-part meronomic relationship, in this research we will look at the ratio between the whole-word frequency and the frequency of the base word and we will carry out a masked priming experiment associated with a lexical decision task, in order to examine suffixed word processing in L1 vs L2.

The poster will describe: a) the research hypothesis, b) the research methodology (i.e. the masked-prime paradigm, Forster & Davis, 1984), c) the choice of L1 and L2 subjects for our experiment, d) the critical materials designed for the research, e) our predictions about experimental results.

More particularly, about the materials, we will focus on: i) the choice of suffixes according to their numerosity and frequency (and productivity/numerosity ratio); ii) the choice of the suffixed words in relation to the relation between the frequency of the whole word and the frequency of the base (suffixed/base ratio), as in the following table:

Derived word	Frequency	Range	Base word	Frequency	Range	Derived/ base ratio
GIOCOSO	13	12383	GIOCO	938	297	72,15385
SETTORIALE	7	18333	SETTORE	493	695	70,4285714
PRESTIGIOSO	104	2873	PRESTIGIO	75	3481	0,721154
PEDONALE	20	9742	PEDONE	11	13653	0,55

References

- Beauvillain C. (1996) The integration of morphological and whole-word form information during eye-fixation on prefixed and suffixed words. *Journal of Memory and Language* 35: 801-820.
- Blumenthal-Dramé A. (2012) *Entrenchment in Usage-Based Theories. What Corpus Data do and do not Reveal about the Mind*. Berlin: de Gruyter Mouton.
- Burani C. & Caramazza A. (1987) The interplay of root, suffix and whole-word frequency in processing derived words. In H. Baayen & R. Schreuder (eds.) *Morphological Structure in Language Processing*, Berlin/New York: Mouton de Gruyter: 157-207.
- Bybee J. (2010) *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

- Colé P., Beauvillain, C. & Segui, J. (1989) On the representation and processing of prefixed and suffixed derived words: A differential frequency effect. *Journal of Memory and Language* 28: 1-13.
- Dal Maso S. & Giraudo H. (in preparation) *The processing of morphologically complex words in Italian L2*. Selected proceedings of the Symposium "Morphology and its interfaces"- Lille, September 12-13, 2013
- Forster K. & Davis C. (1984) Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10: 680-698.
- Tomasello M. (2003) *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press.

Italian Verb-Particle Constructions between structures and usages

Daniela Guglielmo, Università di Salerno

Annibale Elia, Università di Salerno

Azzurra Mancuso, Università di Salerno

This poster focuses on the interface between lexicon, syntax and semantics of a specific set of “conventionalised” structures of Italian, i.e. those formed by a motion verb and a locative adverbial particle such as *andare avanti* (“to go on”), *buttare giù* (“to throw down”), *fare fuori* (lit. to-do-out cf. “to kill”) and so on, called here Verb-Particle Constructions (hereafter VPCs), but also referred to in the literature as phrasal verbs or syntagmatic verbs (Simone, 1997; Cini, 2008, among others). Verb-particle constructions are under the motivation of Construction Grammar (Goldberg, 1995; Jackendoff, 1997; Booij, 2001; Jansen, 2004; Iacobini, 2009) in that they form constructional idioms with an open slot for the verb and a slot that is “preinstalled” with a fixed particle.

The analysis carried out here posits a close relationship between structure facts and usage facts involving Italian VPCs.

The main purpose of this poster is to display the distribution of Italian VPCs in the “concrete” usage of language. In particular, we intend to verify if: (i) these constructions are most frequently employed by language’s users with their prototypical or spatial meaning(s) (e.g. *Max viene fuori di casa* “Max comes out of house”) or rather with their metaphoric or idiomatic meaning(s) (e.g. *Viene fuori un problema* “A problem comes out”) and (ii) if this dichotomy may be satisfactory to account for the large variety of forms and senses disclosed by VPCs in the real experience of the language.

To achieve this goal we perform an automatic extraction of VPCs from Italian texts. The corpus used for this study is the written Italian corpus CoLFIS, containing more than three million words (Bertinetto *et al.*, 1995).

By means of the software Nooj, a new corpus processing system (Silberztein, 2008), we analyze all the occurrences of VPCs (both in terms of types and tokens) resulting from the combination of a motion verb with one of the following locative particles: *avanti* (on), *indietro* (back), *fuori* (out), *dentro* (in/inside), *dietro* (behind/back), *giù* (down), *sopra* (up/on), *sotto* (under/down), *su* (up/on) and *via* (away/off). We extract transitive and intransitive uses, both in continuous and discontinuous patterns.

Then, by means of a lexicon-grammar “splitting of uses” procedure (Gross, 1992; Elia, 2013), which considers in tandem lexicon-syntactic likelihoods and semantic properties, we present the following main families of VPCs uses, as emerged from the corpus:

- 1) Free-compositional uses (e.g. (*correre/andare/venire/...*) *fuori di casa* “(to run/go/come/...) out of the house”);
- 2) Semi-fixed idiomatic uses ((*mettere/sbattere/portare/mandare/tenere*) *dentro il colpevole* “(to put/clap/take/send/keep) the guilty in prison”);
- 3) Fixed idiomatic uses (e.g. *fare fuori il dolce* “to eat up a pastry”);
- 4) Totally fixed idiomatic uses (e.g. *mettere su casa* “to set up home”). In particular, we quantify the frequency distribution of these four structures as well as the frequency distribution of the main spatial and idiomatic meanings disclosed by each particle.

This initial corpus based analysis is supplemented with an experiment of semantic association: forty speakers are asked to produce the first word they think as soon as a specific instance of VPC is presented to them. The following instances of VPCs collected from the corpus are presented to them: (i) the most frequent spatial uses; (ii) the most frequent idiomatic uses; (iii) the less frequent spatial uses; (iv) the less frequent idiomatic uses; (iv) frequency-balanced spatial and idiomatic uses.

The objective is to discover which meaning (spatial vs. idiomatic) of the ambiguous VPCs presented to the speakers (e.g. *buttare giù* “to throw down”, *andare avanti* “to go on”) is more available and ready to be lexically activated by them and if this “fact” is related to the frequency of usage as found in the corpus. In other words, we aim at checking if the results of the experiment confirm the analysis emerged from the corpus exploration.

Factors affecting form-function pairing in vowel perception

Nadja Kerschhofer-Puhalo, Universität Wien

Vowels are complex linguistic structures that vary within and across languages in terms of several articulatory and acoustic phonetic as well as phonological parameters and can be understood as conventionalized pairings of phonetic substance and phonological form. However, there is no one-to-one mapping between the speech signal and the perceived vowel category. Listeners' perception of acoustically different sounds as belonging to the „same“ or a „different“ category shifts as a function of a variety of signal-inherent as well as external factors. Vocalic form-function pairings moreover vary language-specifically; the interpretation of the same acoustic signal may vary with the listeners' native-language background.

Theories on vowel contrasts and vowel systems refer to a general preference for vowel qualities that are maximally distinct from each other and to the assumption that constellations of vowel systems are motivated by maximal contrast, i.e. perceptual distinctiveness (at minimal articulatory cost) (Liljencrants & Lindblom, 1972; Lindblom, 1986; de Boer 2000, 2001; Flemming, 2004; Schwartz *et al.*, 2005). In other words, vowel categories are considered as related to each other by relationships of contrast and distinctiveness or perceptual distance. The *categorization* of sounds is based on perceived identity or similarity. Objectively different vowel sounds are perceived at some cognitive level as more or less similar or distinct from others, belonging either to the same or a different category (Bybee, 2001). *Distance* and *similarity* are gradient concepts, considered to vary along a continuum. An inversely functional correspondence between distance and similarity is assumed in many cognitive models on similarity (e.g. Shepard, 1957; Nosofsky, 1986, 1988). Increasing distance in phonetic substance (form) is considered to increase distinctiveness and guarantees phonemic contrasts (function). By a decrease in distance, on the other hand, which is in inverse functional correspondence to similarity, the form-function correspondence may be at risk. However, no linear correlation of variation in phonetic parameters and perceived phonemic identity can be observed. External factors, such as type of contrast and its use in the listeners' L1 or frequency of occurrence, may contribute to the listeners' reduced capacity to reconstruct the intended form-function pairing in perception.

This contribution addresses the questions of (1) how relationships of physical characteristics and the listeners' perceptions of the speech signal vary language-specifically, (2) how similarity can be operationalized, and (3) what factors may influence the perceived similarity of vowel categories.

Empirical evidence from a vowel identification experiment with L2 learners of German will be presented to illustrate how the interpretation of the same acoustic input varies as a function of signal-inherent factors and external factors. Perceptual confusion data by non-native listeners from different language backgrounds provide evidence for (1) *non-linearities* of changes in phonetic parameters and shifts in phonological categorization, (2) for *language-specific* relationships of similarity and perceptual distance between categories, and (3) for asymmetries and *bias* in vowel categorization.

Based on confusion data and derived similarity scores, the *perceptual vowel space* for German vowels is visualized by means of Multidimensional Scaling (Shepard, 1957, 1980) for each of the ten language sub-groups. The relevance of basic notions of usage-based theory will be discussed to account for the observed relations of perceptual distance and similarity between categories and for the listeners' bias in categorization of vowel sounds in L2 acquisition.

References

- Bybee J. (2001) *Phonology and language use*. New York: Cambridge University Press.
- de Boer B. (2000) Self-organization of vowel systems. *Journal of Phonetics* 28(4): 441-465.
- de Boer B. (2001) *The origins of vowel systems*. Oxford: Oxford University Press.
- Flemming E. (2004) Contrast and perceptual distinctiveness, in: B. Hayes, R. Kirchner & D. Steriade (eds.) *Phonetically Based Phonology*. Cambridge: Cambridge University Press. 232-276.
- Liljencrants J. & Lindblom B. (1972) Numerical simulation of vowel quality systems: the role of perceptual contrast, *Language*, 48: 839-862.
- Lindblom B., Phonetic universals in vowel systems, in: J. Ohala & J. Jaeger (eds), *Experimental phonology*, Orlando, Academic Press: 13-44.
- Nosofsky R. (1991) Stimulus bias, asymmetric similarity, and classification, *Cognitive Psychology* 23: 94-140.

- Nosofsky R. (1986) Attention, similarity, and the identification-categorization relationship, *Journal of Experimental Psychology: General* 115 (1): 39-57.
- Nosofsky R. (1988) Similarity, frequency, and category representations, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(1): 54-65.
- Schwartz J., Abry C., Boë L.J., Ménard, L. & Vallée, N. (2005) Asymmetries in vowel perception in the context of the Dispersion-Focalisation Theory. *Speech Communication*, 45, 425-434.
- Shepard R. (1957) Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space, *Psychometrika* 22: 325-345.
- Shepard R. (1980) Multidimensional Scaling, Tree-Fitting, and Clustering, *Science* 210: 390-398.

Affricate dentali intermedie: emergenza di un nuovo grado di sonorità nell'italiano di Bolzano

Chiara Meluzzi, Libera Università di Bolzano

Il presente lavoro indaga la distribuzione del tratto di sonorità nella pronuncia delle affricate dentali nella varietà di italiano parlata a Bolzano. Per questa indagine si è utilizzato un approccio sociofonetico, con lo scopo di osservare come la micro-variazione fonetica possa assumere valore sociale all'interno di una varietà linguistica relativamente giovane quale quella indagata. La ricerca andrà quindi a osservare la possibilità di evidenziare cambiamenti in atto all'interno di questa varietà assumendo in questo modo una prospettiva *usage-based* (cfr. Bybee, 2001).

Per questo studio è stato raccolto un ampio corpus variabile in diastratia, diafasia e in diacronia apparente: sono stati registrati 42 parlanti italo-foni bolzanini, diversi per età, genere, livello di istruzione e quartiere di residenza, ai quali era richiesta un'intervista sociolinguistica seguita da una lettura di parole. Quest'ultimo task, più formale, ha permesso di elicitarne un totale di 4224 affricate dentali nei diversi contesti fonologici. La variabilità diafasica è stata poi testata su un caso di studio composto da 5 parlanti (3 maschi e 2 femmine) tutti appartenenti allo stesso gruppo familiare, i quali hanno prodotto un totale di 170 affricate dentali nella parte di intervista libera.

L'analisi acustica dei dati ha permesso di evidenziare l'emergenza di alcuni fenomeni peculiari. Tra questi rientra l'emergenza di un tipo di realizzazioni non completamente sonore né completamente sorde: tali foni, definiti come realizzazioni intermedie, presentano la barra di sonorità solo nel segmento occlusivo dell'affricata dentale, laddove la sonorità scompare nel successivo segmento fricativo. Le successive analisi statistiche condotte tramite il software SPSS hanno permesso di rilevare come le produzioni intermedie rappresentino un fenomeno sistematico all'interno dell'italiano di Bolzano. In particolare, è stato possibile osservare come tali foni ricorressero principalmente nel contesto post-sonorantico (es. calzino) e intervocalico (es. dazi-pazzi), oltre a caratterizzare le produzioni dei parlanti più anziani, delle donne e dei parlanti di basso livello di istruzione. Questo pattern di variazione viene confermato anche a livello diafasico nel nostro caso di studio, in cui si evidenzia come le realizzazioni intermedie, quasi assenti nell'eloquio più controllato dei soggetti maschili di giovane età, siano invece presenti in frequenze anche elevate in un eloquio più spontaneo, come emerge dall'intervista libera.

La distribuzione delle produzioni intermedie nel nostro corpus può essere indicativa dell'origine di questo tipo di realizzazioni, originatesi probabilmente per contatto tra le diverse varietà di italiano regionale e i diversi dialetti romanzi che costituiscono la biografia linguistica della comunità italo-fona, quasi come una variabile di accomodamento. Inoltre, si può osservare come questo tratto abbia assunto valore sociale nel sistema dell'italofonia bolzanina: dal momento che si riscontra in maniera elevata nel parlato spontaneo dei più giovani, le intermedie sembrerebbero essere avvertite come tratto identitario, anche se probabilmente molto marcato e pertanto quasi del tutto abbandonato nel parlato più controllato dei soggetti più giovani e istruiti.

L'analisi della micro-variazione fonetica ha dunque evidenziato come ci siano state modificazioni nell'uso di alcuni elementi linguistici che, assumendo valore sociale, hanno contribuito o stanno contribuendo attivamente alla ridefinizione dell'italiano di Bolzano come varietà linguistica autonoma.

Riferimenti

- Bybee J. (2001) *Phonology and Language Use*, Cambridge: Cambridge Studies in Linguistics.
Foulkes P. & Docherty P. (2006) The social life of phonetics and phonology, *Journal of Phonetics* 34: 409-438.
Kerswill P. & Williams A. (2005) New towns and koineisation: linguistic and social correlates *Linguistics* 43(5): 1023-1048.
Meluzzi C. (2013) *Le affricate dentali nell'italiano di Bolzano. Un approccio sociofonetico*, Università di Pavia/Libera Università di Bolzano: Tesi di dottorato.

Frequency distribution of inflectional classes and grammatical gender: a lexical database for Italian nouns.

Francesca Postiglione, *Fondazione Marica De Vincenzi ONLUS*
Maria De Martino, *Università di Salerno*
Giulia Bracco, *Università di Salerno*
Alessandro Laudanna, *Università di Salerno*

Grammatical gender has been defined as a “nominal agreement class” (Corbett, 1991) since it allows to set apart different classes of nouns in many languages. In Italian, gender distinctions are often accomplished on morphological grounds and they are reflected in the orthography/phonology of nouns, where different inflectional paradigms, different degrees of formal regularities and different degrees of gender-related transparency are displayed.

Empirical data demonstrate that speakers are sensitive to information about grammatical gender of words along several dimensions (Gollan & Frost, 2001; De Martino *et al.*, 2011; Paolieri *et al.*, 2011). It has been hypothesised that lexical access reflects the distributional patterns of different grammatical genders and different inflectional paradigms of nouns.

In this work, we describe a lexical database set up in order to obtain distributional data about gender and declensional classes of Italian nouns.

From the CoLFIS database (about 4.000.000 occurrences, Bertinetto *et al.*, 2005), types and tokens corresponding to nouns were extracted, and quantitative and qualitative analyses were performed. For each type and token, starting from the lexical information provided by the CoLFIS database (frequency, length, grammatical class), we provided information about gender, number and inflectional class. Semantic information was not considered (i.e., epicenes). The tagging operations allowed us to estimate the numerosity and the cumulative frequency of gender types and inflectional patterns.

At present, the database is the widest and more representative source of information on gender and inflectional classes for Italian nouns. The database can be exploited for future research and to further verify theoretical explanations of existing data about the role of syntactic (gender), formal (transparency of gender markers) and morphological (inflectional class) features of nouns during lexical processing.

References

- Bertinetto P., Burani C., Laudanna A., Marconi L., Ratti D., Rolando C. & Thornton A.M. (2005) Corpus e Lessico di frequenza dell'Italiano Scritto (CoLFIS). www.ge.ilc.cnr.it/strumenti.php.
- Corbett G. (1991) *Gender*, Cambridge: Cambridge University Press.
- De Martino M., Bracco G. & Laudanna A. (2011) The activation of grammatical gender information in processing Italian nouns, *Language and Cognitive Processes* 26: 745-776.
- Gollan T. H. & Frost R. (2001), Two routes to grammatical gender: Evidence from Hebrew, *Journal of Psycholinguistic Research* 30: 627-651.
- Paolieri D., Lotto L., Leoncini D., Cubelli R. & Job R. (2011) Differential effects of grammatical gender and gender inflection in bare noun production, *British Journal of Psychology* 102: 19-36.

The Statistics of Obsolescence: Case Studies from Late Modern English Grammar. A pilot corpus study

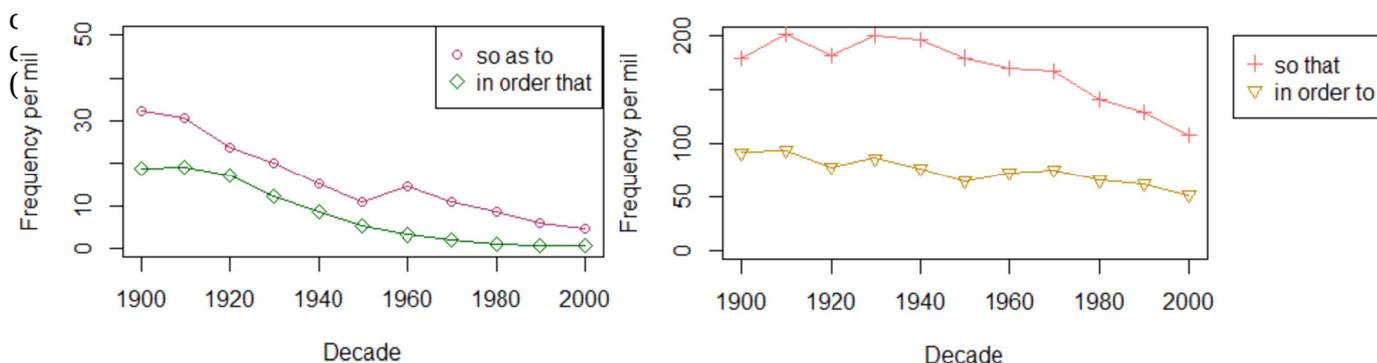
Karolina Rudnicka, Albert-Ludwigs-Universität Freiburg
Graduiertenkolleg DFG GRK 1624 "Frequency Effects in Language"

Purpose: The main goal of the project is to analyze and describe processes of grammatical decline and obsolescence in the recent history of English (ca. 1750 to the present) in the framework of usage-based approaches. In addition to straightforward loss of constructions or fossilization, I also study cases in which decrease in usage frequency leads to reconfigurations in the periphery of the grammatical system. The pilot study that is going to be presented at the poster is devoted to corpus-based statistical observations of the occurrence of 4 constructions relating to purpose, namely: *in order that*, *in order to*, *so that* and *so as to*.

Methods: The methods that were used in the pilot study are based on corpus investigation and application of correlation testing in the R software. The corpora explored in the first place were the Corpus of Historical American English (COHA) and the Corpus of Contemporary American English (COCA). Obtained results are further interpreted in terms of the usage-based approach.

Results: The results of the searches have shown there is a quite visible decline in the frequency of occurrence of the constructions *in order that* and *so that*. In the case of *in order to* and *so as to* the decline seems to take a little less steep form but it is still present.

course is not a mere fluctuation is further



Conclusion: The negative correlation measured by Kendall's tau test can be according to Gries (2013, p. 147, table 18) assessed as being very high in the case of all four candidate features. The above trend towards the decline of some constructions expressing purpose is now being discussed in the framework of the usage-based approach.

Pragmatic markers from a construction grammar perspective: Italian *alla fine* and German *im Endeffekt*

Valentina Russo, Humboldt-Universität zu Berlin | Università di Napoli “L’Orientale”

This poster aims at describing a work in progress analysis of two polyfunctional linguistic expressions used in spoken Italian and German, i.e. *alla fine* and *im Endeffekt*. These two fixed formulas are entrenched as idiomatic units with non-compositional meaning, which are able to fulfil many pragmatic functions in different constructions, ranging from modal to discursive ones. Taking a construction grammatical perspective, this study presents an analysis of the formal and functional properties of the mentioned items in form-meaning pairings, i.e. *constructions* (s. Goldberg, 1995; Croft, 2001). For similar expressions in other European languages (such as en. *after all*, *besides*, *indeed*; fr. *après tout*, *en somme*, *enfin*, etc.) it has been claimed they may act as markers for (among others): conclusion, hesitation, modality, aspect, inference, re-examination, reformulation, subjectivity, argumentation, afterthoughts, repairs (s. Philippi, 1999 for an overview of works mainly based on modular approaches). As well known, the boundaries between such categories, in particular between modal and discursive particles, are rather fuzzy (each showing overlapping functions with respect to one another), and different categorizations mostly depend on different approaches to the same subject (s. Degand *et al.*, 2013).

In this paper I suggest to describe and compare Italian *alla fine* and German *im Endeffekt* in a cross-linguistic perspective through a corpus-based analysis by adopting a construction grammatical approach. Indeed, Construction Grammar presents many advantages for such an analysis since:

1) As a usage-based approach, it enables a fine-grained analysis of “less formal” expressions, especially those mainly used in oral discourse, understanding constructions as “psychologically entrenched and conventional in the speech community” (Langacker, 2005: 140);

2) As a non-modular approach, it takes into account different levels of analysis (ranging from morpho-syntactic to sentence- or higher-level sequences) considering them as a continuum;

3) It has already proved to be suitable for discourse and conversational analysis (s. Auer, 2006; Fischer, 2008; Östman, 2005) as well as for categorizing modal and discursive particles – especially in heteroseme cases (s. Imo, 2008; Diewald, 2013; Fischer & Alm, 2013) – and for dealing with language-specific constraints (s. Dittmar, 2012; Russo & Dittmar, *in press.*).

Given the *No-Synonymy Principle* formulated by Goldberg (1995), my analysis will, thus, account for specific features of uses of *alla fine* and *im Endeffekt* that contribute to the instantiation of different constructions (i.e. syntactic position and integration at utterance- and discourse-level; scope, semantics and pragmatics; prosodic profile and integration in the main utterance; context and communicative background; co-occurrence with specific co-selected items), showing once again how the notion of grammatical construction is a useful analytical tool for accounting for the interaction of syntactic, semantic, and phonological factors.

Finally, I will try to sketch a network of related constructions sharing aspects of form and function (s. Goldberg & van der Auwera, 2012), linking the various senses and uses of *alla fine* and *im Endeffekt* with each other.

References

- Auer P. (2006) Construction Grammar meets Conversation: Einige Überlegungen am Beispiel von ,so’-Konstruktionen, in G. Suzanne & I. Wolfgang (eds.), *Konstruktionen in der Interaktion*, Berlin: De Gruyter: 291-314.
- Croft W. (2001) *Radical Construction Grammar*, Oxford: Oxford University Press.
- Degand L., Cornillie B. & Pietrandrea P. (2013), Modal particles and discourse markers: Two sides of the same coin?”, in Degand *et al.* (eds.) *Discourse Markers and Modal Particles: Categorization and description*, Pragmatics and Beyond New Series, v. 234, Amsterdam, John Benjamin Publishing Co., 1-18.
- Diewald G. (2013) ‘Same same but different’ – Modal particles, discourse markers and the art (and purpose) of categorization, in Degand *et al.* (eds.), *Discourse Markers and Modal Particles: Categorization and description* 19-45.
- Dittmar N. (2012) Costruire il parlato: macro- e micro-sintassi“, in F. Orletti, E. Lombardi Vallauri & A. Pompei (Hgg.) *Grammatica e pragmatica. Atti del Convegno SIG*, Roma Il Calamo: 87-118.

- Fischer K. (2008) Konstruktionsgrammatik und Interaktion, in K. Fischer & A. Stefanowitsch (eds.), *Konstruktionsgrammatik I. Von der Anwendung zur Theorie*, Tübingen, Stauffenburg: 133-150.
- Fischer K. & Alm M. (2013) A radical construction grammar perspective on the modal particle-discourse particle distinction, in Degand et al. (eds.), *Discourse Markers and Modal Particles: Categorization and description*: 47-87.
- Goldberg A (1995), *Constructions. A Construction Grammar Approach to Argument Structure*, Chicago, The University of Chicago Press.
- Goldberg J. & van der Auwera J. (2012) This is to count as a construction, in *Folia Linguistica* 46: 109-132.
- Imo W. (2008) Individuelle Konstrukte oder Vorboten einer neuen Konstruktion? Stellungsvarianten der Modalpartikel *halt* im Vor- und Nachfeld, in A. Stefanowitsch / K. Fischer K. (eds.), *Konstruktionsgrammatik II. Von der Konstruktion zu Grammatik*, Tübingen: Stauffenburg Verlag, 135-156.
- Langacker R. (2005) Construction grammars: Cognitive, radical, and less so, in F. Ruiz de Mendoza Ibanez & M. Sandra Pena Cervel (eds.) *Cognitive linguistics: Internal dynamics and interdisciplinary interaction*, Berlin, de Gruyter: 101-159.
- Östman J. (2005) Construction discourse: a prolegomenon, in: J. Östman & M. Fried (eds.), *Construction Grammars. Cognitive Grounding and Theoretical Extensions*, John Benjamins, Amsterdam: 121-144.
- Philippi D. (1999) *Après tout als pragmatischer Indikator im Französischen*, RomLi/11, Berlin Freie Universität Berlin.
- Russo V. & Dittmar N. (*in press*) Konstruktionen konversationeller Vagheit in deutschen und italienischen Gesprächen, in: Selig, Dittmar & Morlicchio (eds.), *Gesprächsanalyse zwischen Syntax und Pragmatik. Deutsche und italienische Konstruktionen*, Tübingen: Stauffenburg Verlag.

Judgements of collocation frequency in L1 and L2: Intuition vs. evidence from corpora

Anna Siyanova-Chanturia, Victoria University of Wellington
Stefania Spina, Università per Stranieri di Perugia

Native and non-native speaker intuitions about collocation frequency are a largely unexplored area (Siyanova-Chanturia *et al.*, 2011): to date, research has focused primarily on native and non-native speaker intuitions about *single word* frequency (for native speakers: Alderson, 2007; Backman, 1976; Carroll, 1971; Frey, 1981; McCrostie, 2007; Shapiro, 1969; Tryk, 1968; for non-native speakers: Ringeling, 1984; Schmitt & Dunham, 1999). To bridge this gap, the present study aimed to answer the following questions:

- How do L2 learners and native speakers compare against evidence from corpora in their subjective judgements of collocation frequency?
- How do L2 learners and native speakers compare against each other in their subjective judgements of collocation frequency?

84 participants (42 native speakers, 21 advanced and 21 intermediate learners of Italian) were asked to judge 80 noun+adjective collocations as one of the following: high frequency (*tempo libero* “free time”), medium frequency (*guida turistica* “tourist guide”), low frequency (*difetto fisico* “physical defect”), or very low frequency (*nonni ospitali* “hospitable grandparents”). Frequencies were extracted from two Italian multi-genre corpora: *Perugia* corpus (<http://perugiacorpus.unistrapg.it/>) and *La Repubblica* corpus (Baroni *et al.*, 2004).

Two types of analyses were conducted. First, Cohen’s kappa statistic (Cohen, 1960) revealed that both native speaker and L2 learner intuitions of high frequency collocations correlated strongly with corpus frequencies. Neither of the two groups of participants exhibited accurate intuitions - relative to corpus frequencies - of medium and low frequency collocations. With regard to very low frequency items, native speaker but *not* L2 learner intuitions were found to correlate with the corpora for the majority of the items.

Second, mixed effects modelling (Baayen *et al.*, 2008) revealed that both L2 learners and native speakers were sensitive to the frequency of items within the four frequency bands, and that the level of proficiency played a role and interacted with other factors. Specifically, native speaker, advanced and intermediate learner judgements of collocation frequency were influenced differently by the length of the first word of the collocation.

In sum, the present study suggests that native speakers:

- possess better intuitions of collocation frequency than non-native speakers in the case of very low frequency items;
- possess accurate intuitions of collocation frequency only in the case of the extreme frequency groups (very high and very low).

Taken together, the study provides new insights into the nature of L1 and L2 intuition about phrasal frequency, as well as into the role of frequency of exposure: if our mental representations of linguistic items are determined by language use, in other words, by *frequency of occurrence* (Bybee, 2006; Bybee & Hopper, 2001; Ellis, 2012; Goldberg, 2006; Tomasello, 2003), then this different way of developing frequency intuitions has to be ascribed to frequency of exposure. Native speakers with a sufficient amount of exposure to Italian noun+adjective collocations of various frequencies exhibited better intuitions about very low frequency items than did non-native speakers, whose exposure to Italian has not been as rich.

References

- Alderson J. (2007) Judging the frequency of English words, *Applied Linguistics* 28(3): 383-409.
- Baayen R., Davidson D. & Bates D. (2008) Mixed-effects modeling with crossed random effects for subjects and items, *Journal of Memory and Language* 59: 390-412.
- Backman J. (1976) Some common word attributes and their relations to objective frequency counts, *Scandinavian Journal of Educational Research* 20: 175-186.
- Baroni M. *et al.* (2004) Introducing the La Repubblica corpus: a large, annotated, TEI (XML)-compliant corpus of newspaper Italian, in M. Lino (ed.), *Proceedings of the Fourth International Conference on Language Resources and Evaluation 2*, Paris, ELRA - European Language Resources: 1771-1774.
- Bybee, J. (2006) From usage to grammar: The mind’s response to repetition, *Language* 82(4): 711-733.

- Bybee J. & Hopper P. (2001) *Frequency and the Emergence of Linguistic Structure*, Amsterdam, Benjamins.
- Carroll J. B. (1971) Measurement properties of subjective magnitude estimates of word frequency, *Journal of Verbal Learning and Verbal Behavior* 10: 722-729.
- Cohen J. (1960) A coefficient of agreement for nominal scales, *Educational and Psychological Measurement* 20(1): 37-46.
- Ellis N. (2012) What can we count in language, and what counts in language acquisition, cognition, and use?, in S. Gries & D. S. Divjak (eds.) *Frequency effects in language learning and processing (Vol. 1)*, Berlin, Mouton de Gruyter: 7-34.
- Frey E. (1981) Subjective word frequency estimates and their stylistic relevance in literature, *Poetics* 10(4-5): 395-407.
- Goldberg A. (2006) *Constructions at Work: The Nature of Generalization in Language*, Oxford, Oxford University Press.
- McCrostie J. (2007) Investigating the Accuracy of Teachers' Word Frequency Intuitions, *RELC Journal*, 38(1): 53-66.
- Ringeling T. (1984) Subjective estimates as a useful alternative to word frequency counts, *Interlanguage Studies Bulletin*, Utrecht, 8(1): 59-69.
- Schmitt N. & Dunham B. (1999) Exploring native and non-native intuitions of word frequency, *Second Language Research* 15(4): 389-411.
- Shapiro B. J. (1969) The subjective estimation of relative word frequency, *Journal of Verbal Learning and Verbal Behavior* 8: 248-251.
- Sivanova-Chanturia A., Conklin K. & van Heuven W. (2011) Seeing a phrase 'time and again' matters: The role of phrasal frequency in the processing of multi-word sequences, *Journal of Experimental Psychology: Language, Memory, and Cognition* 37(3): 776-784.
- Tomasello M. (2003) *Constructing a Language: A Usage-Based Theory of Language Acquisition*, Cambridge, London: Harvard University Press.
- Tryk H. E. (1968) Subjective scaling of word frequency, *American Journal of Psychology*, 81: 170-177.

Does frequency win over syntactic complexity? Cross-linguistic evidence from the processing of Relative vs. Noun-Clausal Complement structures

Mirta Vernice, *Università di Milano-Bicocca*

Vincenzo Moscati, *Macquarie University*

Caterina Donati, *Università di Roma La Sapienza*

Carlo Cecchetto, *Università di Milano-Bicocca*

The study of complex sentences is pivotal for our understanding of the various processing factors that influence the human parser. We will present here a comparative study on Italian and English. By using an on-line methodology, we focused on the real-time processing of Relative Clauses (RC) in order to assess the role played by frequency and other syntactic factors (see Donati & Cecchetto, 2011) in the resolution of temporal ambiguities.

Relative clauses (RC) are complex, since they contain a filler-gap dependency, with a well-known modulation (object RCs harder than subject RCs). A *that*-relative like ‘the claim that John made (is false)’ is temporarily ambiguous with a Noun Clausal-Complement (CC) structure like ‘the claim that John made (a mistake)’. Since CCs contain no gap, they should be simpler. However, CCs are much more frequent. So, RCs and CCs are the ideal situation to contrast effects of frequency and of syntactic complexity.

In a first eye-tracking experiment, we asked 28 Italian participants to read 24 pairs of object-RCs and their corresponding CCs (1) and 24 sets of subject-RCs and their corresponding CCs (2). Each pair of sentences differed only in critical area 3. In addition, we estimated the effect of three sources of frequency: the clausal structure preferences of each noun (based on a internet-based search), the preferred completions for sentence fragments for each noun (based on a completion task) and the relative frequency of the verb form used (subjunctive versus indicative).

(1)¹Il progetto che ²il sindacato ³sosteneva fra vs. sostenesse ⁴gli operai ⁵irritò la Confindustria/

“The plan that the Trade Unions supported among/would support the workers annoyed the entrepreneurs”

(2)¹Come sempre ²l’insinuazione che ³creava vs. creassero ⁴scandalo ⁵veniva dai giornali/

“As usual, the insinuation that fabricated/ that (they) should fabricate a scandal came from the press”

Regression measures at area 3 indicated that CCs required re-analysis in comparison with object-RCs. Accordingly, in the same area, CCs caused longer reading times as compared to subject-RCs. Overall, results indicate that participants initially opt for the RC interpretation. The analysis of the frequency variables revealed that the less frequent the verb form (i.e., subjunctive) the longer the reading times. Therefore, as in Italian CCs the verb appears mostly in the subjunctive form, whereas RCs involve only the indicative, one could not exclude that our results are an artifact of the verb form frequency bias.

To disentangle this issue, we conducted a second experiment in English (24 participants), replicating Experiment 1. We created 24 sets of object-RCs and their corresponding CCs and other 24 sets of subject-RCs and CCs pairs. Additionally we estimated the clausal structure preferences for each noun.

(3)¹The conclusion that ²they had reached ³the end vs. at the end ⁴of their work ⁵elicited warm applause./

(4)¹Apparently ²the fact that ³upsets vs. he upsets ⁴me seems to be ⁵completely irrelevant to you./

The results showed a pattern comparable to the Italian experiment: CCs elicited more regressions at area 4 as compared to object-RCs; again, at area 4, CCs slowed down reading times in comparison with subject-RCs. In contrast, the frequency variable showed a minor predictive effect.

We discuss the consistency of the pattern of results found in both languages with respect to a syntactic-based analysis of relatives (cf. Donati & Cecchetto, 2011).

Small number constructions: intentional vagueness at work

Miriam Voghera, Università degli Studi di Salerno

We present here a usage-based account of a set of constructions in Italian, French and Spanish in which small numbers are used to express paucal meaning:

- (1) J'ai écrit **deux lignes**
Lit. I have written two lines 'I wrote a few words'
- (2) Ho messo **quattro cose** in valigia
Lit. Have.I pers. put four things in suitcase 'I put a few things in my suitcase'
- (3) Dicho en **dos palabras**
Lit. Said in two words 'Said in few words'

In these constructions small numbers lose their cardinal meaning to express an approximate quantity, as paucal normally does (Corbett, 2000). The paucal interpretation of small numbers depends on both cognitive and linguistic conditions. Cognitively "[...] human infants exhibit [...] approximate number discrimination and comparison abilities in the first year of life, before the acquisition of number words" (Lemer *et al.* 2003: 1942). This allows for the capacity of estimate quantity without the necessity of expressing a precise measure. Linguistically, the vagueness expressed by small number is just one of the signs of a basic intrinsic property of human language, which manifests itself at all level of codification and in a number of different ways (Russell 1923, Wittgenstein 1953, De Mauro 1982, Channel 1994). In fact, small number constructions (SNC) are part of a wider family of constructions in which numbers are used to express intentional vagueness (Voghera 2012), such as vague quantifiers, vague amounts and numbers, paucal, multal quantifiers, hyperbolic numerals (Channel 2004; Krifka 2007; Lavric 2010; Bazzanella 2011).

Using data from different spoken corpora, we found that SNCs manifest restrictions of different nature: paradigmatic restrictions on kinds of numbers and nouns that can be used, depending on roundness and granularity (Devos *et al.* 1998; Jansen & Pollmann 2001); syntactic restrictions on the constituency of the NP they can form (\pm det/ \pm adj); pragmatic restrictions depending on the kind of situation and relation between the speakers participating to the communicative exchange. Not all the restrictions apply to all kinds of SNCs: for instance, coordinate numbers, such as *due o tre* 'two or three', present less restrictions compared with the use of single numbers. In fact, the combinations of different restrictions design a continuum of SNCs representing different degrees of cohesion, fixedness and idiomaticity, which in turn may or may not correlate with paucal interpretation. Although the vague interpretation of small numbers is theoretically always possible, our work shows that it is actually related to the form of construction in which the small numbers occur.

References

- Bazzanella C. (2011) *Numeri per parlare. Da 'quattro chiacchiere' a 'grazie mille'*, Roma: Laterza.
- Channell J. (1994) *Vague language*, Oxford: Oxford University Press.
- Corbett G. (2000) *Number*, Cambridge: Cambridge University Press.
- De Mauro T. (1982) *Minisemantica dei linguaggi non verbali e delle lingue*, Roma: Laterza.
- Devos F., Maesfranckx, P. & De Tré, G. (1998) Granularity in the interpretation of around in approximative lexical time indication, *Journal of Quantitative Linguistics* 5(3): 167-173.
- Krifka M. (2007) Approximate interpretation of number words: A case for strategic communication, *Cognitive foundations of interpretation*: 111-126.
- Lavric E. (2010) Hyperbolic approximative numerals in cross-cultural comparison, in G. Kaltenböck, W. Mihatsch & S. Schneider (eds.) *New Approaches to Hedging*, Bingley: Emerald Group Publishing Limited. 123-164.
- Lemer C. *et al.*, (2003) Approximate quantities and exact number words: dissociable systems, *Neuropsychologia* 41: 1942-1958.
- Russell B. (1923) Vagueness, *The Australasian Journal of Psychology and Philosophy* 1: 84-92.
- Voghera M. (2012) Chitarre, violini, banjo e cose del genere, in: A. Thornton & M. Voghera (a cura di) *Per Tullio De Mauro. Studi offerti dalle allieve in occasione del suo 80° compleanno*, Rom: Aracne: 341-364.

Wittgenstein L. (1953) *Philosophische Untersuchungen*.

Pragmatic specialization of transitive *be* perfects in Canadian English: a corpus study

Yuri Yerastov, Fort Hays State University

I present a corpus-based analysis of reflexes of the transitive *be* perfect found in some varieties of North American English – most prominently in Canadian and Philadelphian English – for example: *I am done dinner, I am finished homework, I am started my project* (Yerastov, 2012, 2013; Hinnell, 2012, Fruehwald & Myler, 2012). While linguistic scholarship has profusely addressed the grammaticalization of perfects in English, along with functional (Smith 2001, 2007) and sociolinguistic (Ryden, 1991) causes of the *be ~ have* intransitive auxiliary split, it remains unclear what motivates a similar split among transitive verbs in North American dialects. While insular Scottish dialects exhibit fully productive *be* perfects that apply across the board (Millar, 2007; Pavlenko, 1997; Flaws & Lamb, 1996), it is not particularly clear why some North American dialects have entrenched the participles *done, finished, and started* in combination with the auxiliary *be*.

The present study draws on the commercial database *Canadian Newsstand Complete*, comprised of local, regional and national news media sources. Based on the quantitative analysis of over 1,500 exemplars of the construction from this database, I demonstrate that there is a statistically significant preference for definiteness in the determiner slot of the direct object of the construction and for the transitive *be* perfect to occur in preposed adverbial clauses, e.g.

(1)

- a. When you are done your cooking session, turn the bottle off first and let the system drain.
- b. So when you are done your dinner at a restaurant and find your car window smashed, or arrive home from a holiday to find your television is missing, remember as long as the story runs on the bottom of page 4, nothing will be done about it.
- c. If I am doing some work around the house, I will be reminded that when I am finished the project, aesthetics come first.
- d. After the musicians are finished the hockey game, they'll come outside at 3:30 p.m. and perform a concert until 7 p.m.

Canadian Newsstand Complete

I argue that these statistical facts suggest that the *be ~ have* transitive auxiliary split in Canadian English is in part motivated by pragmatic, discourse-based specialization between the *have and be* perfects. The transitive *be* perfect is emerging as a topicalization device; pre-posed adverbial clauses and definiteness markers in the direct object slot are both means to code given, familiar information. I reinforce this argument with a qualitative analysis of contextual cues, which center around a narrow set of conceptual domains such as household duties, athletic activities, and educational achievement.

References

- Flaws M. & Lamb G. (1996) *The Orkney dictionary*, Orkney: Orkney Language and Culture Group.
- Fruehwald J. & Myler N (2013) “*I am done my homework*” – *Case assignment in a stative passive*. University of Pennsylvania linguistics colloquium.
- Hinnell J. (2012) A construction analysis of [be done X] in Canadian English. Master’s thesis. Simon Fraser University.
- Millar R. (2007) *Northern and Insular Scots*. Edinburgh: Edinburgh University Press.
- Pavlenko A. (1997) The origin of the *be*-perfect with transitives in the Shetland dialect, *Scottish Language* 16: 88-96.
- Smith K. (2007) Language use and auxiliary selection in the perfect, in ed. R. Aranovich (ed.) *Split auxiliary systems: a cross-linguistic perspective*, Philadelphia/Amsterdam: Benjamins: 255-270.
- Smith K. (2001) The role of frequency in the specialization of the English anterior, in J. Bybee (ed.) *Frequency and the emergence of linguistic structure*, Amsterdam, Benjamins: 361- 382.
- Rydén, Mats. 1991. The *be/ have* variation with intransitives in its crucial phrases, in D. Kastovsky (ed.) *Historical English Syntax*, Berlin, De Gruyter: 343-354.
- Yerastov Y. (2013) A synchronic analysis of the transitive *be* perfect in Canadian English.” *Journal of English Language and Linguistics*. Accepted.
- Yerastov Y. (2012) Transitive *be* perfect in Canada: An experimental study, *Canadian Journal of Linguistics* 57(3): 1001-1031.

