

IØRN KORZEN

## Cosa ci rivelano i corpora sulla complessità testuale dell'italiano?

In questo lavoro discuto la nozione di complessità linguistica con particolare riguardo alla strutturazione testuale. Paragono l'italiano e il danese e come base empirica mi servo di tre corpora di testi paralleli di tipologie diverse: il *corpus Europarl* (discorsi argomentativi tenutisi al Parlamento europeo), il *corpus di Mr. Bean* (90 esposizioni narrative, scritte e orali, di due episodi di Mr. Bean prodotte da studenti universitari di Torino e di Copenaghen) e il *corpus SugarTexts* (testi espositivi tecnici sulla produzione di zucchero da barbabietola). In questi testi analizzo due fenomeni che si differenziano cross-linguisticamente in modo particolare: il numero di proposizioni per periodo e la loro testualizzazione finita vs. non-finita, ossia il grado della loro deverbizzazione. Con riferimento a questi due fenomeni non vi è dubbio che l'italiano si manifesti come lingua assai più complessa del danese.

*Parole chiave:* complessità linguistica, strutturazione testuale, corpora paralleli, densità testuale, deverbizzazione.

### 1. Introduzione

“L'italiano è una lingua complessa e difficile!”. Sono più di 40 anni che sento regolarmente, in qualità di docente di italiano in Danimarca, tale lamentela da parte dei miei studenti, e colleghi docenti delle altre lingue romanze sentono lagnanze simili. Ma è davvero così? Alcune lingue o gruppi linguistici sono veramente più complessi di altri? E se sì, secondo quali parametri, a quali livelli linguistici e, inoltre, come verificare tale ipotesi?

La nozione di “complessità” è approdata alla linguistica da altre scienze quali fisica, antropologia e filosofia, ed è generalmente definita come la quantità di informazioni necessarie per comprendere e spiegare adeguatamente un sistema o un insieme di elementi (Merlini Barbaresi 2003: 24, 2005: 302). In linguistica il concetto di complessità è stato ampiamente indagato e discusso negli ultimi decenni, spes-

so con definizioni simili. Alcuni studiosi distinguono tra complessità “assoluta”, ossia intrinseca al sistema linguistico stesso, e complessità “relativa”, ossia vissuta soggettivamente dai parlanti e basata su specifici aspetti problematici per parlanti e/o apprendenti<sup>1</sup>.

Nel suo approccio sociolinguistico, Moretti (2018: 40-42) distingue tra complessità “spontanea o primaria”, legata alla creazione della lingua in questione, e complessità “secondaria”, causata ad esempio dalla normativizzazione di varietà particolarmente formali, caratteristica molto evidente dell’italiano e confermata anche da Berruto (2012: 176) che – in accordo con i miei studenti – conclude: “Non converrebbe, in conclusione, sostenere che, più che essere l’italiano popolare una varietà linguistica semplificata, sia l’italiano standard e colto una varietà linguistica particolarmente complessa, elaborata, in un certo senso ‘innaturale’, grazie appunto alla sua precoce standardizzazione letteraria, aulica e elitaria...?”.

Le pagine seguenti sono strutturate in questo modo: nella sezione 2 citerò altri punti di vista e considerazioni degli studiosi sul concetto di complessità, dopodiché, nelle sezioni 3-4, presenterò i corpora di testi paralleli sui quali baserò le mie analisi e osservazioni: il *corpus Europarl*, il *corpus di Mr. Bean* e il *corpus SugarTexts*. Paragonerò la lunghezza dei periodi italiani e danesi misurata come numero di parole e di proposizioni per periodo, e nella sezione 5 confronterò la testualizzazione (esplicita, implicita o nominalizzata) delle singole proposizioni dei tre corpora. Sulla base dei miei risultati tirerò alcune conclusioni nella sezione 6.

## 2. Definizioni e prime indicazioni di diversa complessità italiana-danese

Probabilmente è impossibile definire la complessità complessiva di una lingua in modo oggettivo e razionale (Deutscher 2009: 247); nella migliore delle ipotesi, essa può essere intesa come l’accumulo di diversi valori che insieme accrescono o diminuiscono la complessità di particolari dimensioni linguistiche. Va aggiunto che la complessità “relativa” e “learner-related” non è mai costante dato che dipende lar-

<sup>1</sup> Cfr. per esempio Masi (2003), Merlini Barbaresi (2004: sez. 1), Miestamo (2009: 81ff), Dahl (2009: 50-52) e Fiorentino (2009: 282ff).

gamente dalla distanza tra la L1 dell'apprendente e la L2 in questione<sup>2</sup>. Anche la nozione di complessità testuale risulta alquanto complessa: “More precisely, textual complexity turns out to be the result of the cumulative effects of the interaction among different variables that belong to all the levels of texture.” (Masi 2003: 142).

Molti studiosi che propongono definizioni più specifiche adottano come punto di partenza nell'approccio generale alla complessità di un oggetto “the amount of information needed to recreate or specify it” (Dahl 2009: 50). McWhorter (2001: 125) definisce la complessità linguistica come “the degree of overt signalling of various phonetic, morphological, syntactic, and semantic distinctions beyond communicative necessity”, e come esempi egli menziona marcatura di genere, multipli tempi passati, il congiuntivo e verbi nominalizzati. Nello stesso spirito, Fiorentino (2009: 282-286) cita, tra i fattori che contribuiscono alla complessità linguistica, un alto numero di sottotipi o di varianti alternative di un dato elemento o di una data funzione (per esempio morfologica), un alto numero di regole sintattiche e poca trasparenza nella relazione tra forma e funzione. Per una definizione simile, vedi Nichols (2009: 111-112).

In un'analisi delle differenze tra lingua parlata e scritta, Chafe (1985) introduce il concetto di “idea units”, ossia “unità di pensiero”, che contengono “all the information a speaker can handle in a single focus of consciousness” (Chafe 1985: 106). Nella lingua scritta tali unità di pensiero sono testualizzate come periodi, e chiare indicazioni di differenze di testualizzazione e di complessità, per esempio tra l'italiano e il danese, possono essere ottenute da corpora multilingui di testi paralleli, cioè testi autentici prodotti indipendentemente nelle lingue in questione, ma in situazioni equivalenti e per target e con contenuti equivalenti.

---

<sup>2</sup> Cfr. Deutscher (loc.cit), Bertuccelli (2003: 139), Maas (2009: 177), Nichols (2009: 120ff), Moretti (2018: 37), Korzen (2021). Sul tema dell'acquisizione della seconda lingua (SLA), cfr. anche Ellis (2016).

### 3. Il corpus “Europarl”

Un corpus frequentemente adoperato per paragoni linguistici è il *corpus Europarl*, e un paragone per esempio dei periodi di tutti i testi italiani e danesi L1 degli anni 1996-2010 porta ai seguenti risultati:

Tabella 1 - *Corpus Europarl: lunghezza dei periodi*

<i>Corpus Europarl</i> (anni 1996-2010)	1. Parole <sup>3</sup>	2. Periodi	3. Parole per periodo	
			Numeri medi	Differenza
<i>Testi italiani</i>	1.657.592	47.405	35,0	45,2 %
<i>Testi danesi</i>	546.425	22.668	24,1	

Come risulta dalla Tabella 1, misurati come numero medio di parole, i periodi italiani sono notevolmente più lunghi di quelli danesi, la differenza arrivando al 45,2 %<sup>4</sup>. Però di per sé, il semplice numero di parole per frase non è necessariamente indicativo di una complessità alta o bassa (Masi 2003: 141; Merlini Barbaresi 2004: sez. 2; Fiorentino 2009: 309; Korzen 2021: 22): bisogna capire a che cosa servono le molte “parole in più” nei testi italiani, se semplicemente a descrizioni più specificate e dettagliate oppure ad un numero più elevato di fatti e avvenimenti.

A tal fine, ho calcolato e paragonato il numero di proposizioni di tutti i periodi, ossia il numero di unità consistenti di un predicato e relativi argomenti, unità che possono essere testualizzate come frasi principali o subordinate, e se subordinate: esplicite, implicite o nominalizzate. Dato che tutti questi calcoli e paragoni sono stati condotti manualmente, qui mi sono servito di un sottocorpus *Europarl* di dimensioni più modeste, più precisamente di 70 testi in ogni lingua, scel-

<sup>3</sup> Le differenze di rappresentanza italiana e danese nel Parlamento europeo rendono i testi italiani degli anni in questione circa tre volte più numerosi dei testi danesi.

<sup>4</sup> Naturalmente, varie riserve vanno fatte nell’operare con calcoli basati sull’unità “parola (grafica)”. Alcune differenze tipologiche portano ad un numero di parole più basso in danese (ad esempio l’articolo definito che in danese è spesso enclitico e i molti nomi composti danesi che corrispondono a costrutti nome + preposizione + nome in italiano). Altre differenze comportano invece un numero più basso in italiano (ad esempio verbo + pronome enclitico in italiano, inesistente in danese, e il fenomeno del pro-drop). Tuttavia, la maggior parte delle differenze interlinguistiche menzionate in questa tabella e in quelle seguenti sono di dimensioni che, a mio parere, consentono di utilizzarle come indicazioni di fondamentali differenze testuali.

ti in modo da rappresentare una varietà di riunioni parlamentari, una varietà di argomenti trattati e una varietà di autori; cfr. anche Korzen & Gylling (2017), dove ci eravamo serviti di 50 testi in ogni lingua.

Tabella 2 - *Sottocorpus Europarl: proposizioni e periodi*

<i>Sottocorpus Europarl</i>	1. <i>Testi</i>	2. <i>Parole</i>	3. <i>Proposizioni</i>	4. <i>Periodi</i>	5. <i>Proposizioni per periodo</i>	
					<i>Numeri medi</i>	<i>Differenza</i>
<i>italiano</i>	70	22.707	2.827	685	4,13	57,0 %
<i>danese</i>	70	22.705	2.758	1.049	2,63	

Questa tabella rivela un'altra cosa interessante: i due sottocorpora sono di dimensioni equivalenti quanto al numero di testi e di parole (colonne 1-2) e di dimensioni piuttosto simili quanto al numero di proposizioni (colonna 3). Invece, per quanto riguarda il numero di periodi (colonna 4) si osserva una grande differenza, differenza riscontrabile, di conseguenza, anche nel numero di proposizioni per periodo (colonna 5), dove essa supera perfino la differenza dei numeri di parole per periodo che avevamo visto in Tabella 1, colonna 3. Nei sottocorpora indagati, le parole "in più" nei testi italiani servono quindi a testualizzare nuovi eventi, avvenimenti o situazioni e in quel modo a rendere più concettualmente complessi i singoli periodi.

#### 4. *Altri due corpora*

Come è noto, il *corpus Europarl* consiste di testi argomentativi, più precisamente dei discorsi politici tenutisi al Parlamento europeo (Koehn 2005, <https://statmt.org/europarl/>). Per assicurarmi che le differenze illustrate nelle Tabelle 1-2 non fossero limitate a tale particolare tipo e genere testuale, ho aggiunto alle mie analisi altri due corpora, anch'essi di dimensioni modeste, ma di tipi e generi diversi:

- The *Mr. Bean corpus*, testi narrativi: 90 esposizioni, scritte e orali, di due episodi di Mr. Bean ("The Library" e parte di "Merry Christmas Mr Bean") prodotte da 27 studenti dell'Università di Torino e da 18 studenti dell'Università di Copenaghen. Il corpus fu creato nel 1995 da un gruppo di docenti dell'Università di Copenaghen e della Copenaghen Business School, incluso il sottoscritto (Skytte et al. 1999, <http://blog.cbs.dk/mrbean-korpus/>), con la collaborazione

- dell'Università di Torino e di Carla Bazzanella. Facemmo vedere i due episodi agli studenti chiedendo loro di riferire poi gli episodi oralmente e per iscritto nella propria madrelingua.
- Il corpus *SugarTexts* – *Telling the SugarStory* in diverse languages, (Smith 2009, <http://www.sugartexts.dk/>), che consiste di testi espositivi tecnici sulla produzione di zucchero da barbabietola in sei lingue diverse romanze e germaniche e nel russo. Il corpus focalizza particolarmente i processi effettuati in seguito all'arrivo delle barbabietole agli zuccherifici. I testi sono stati raccolti da dottorandi e laureandi della Copenhagen Business School in collaborazione con i loro tutor, fra cui il sottoscritto, e appartengono a generi svariati come siti web e dépliant di aziende produttrici di zucchero, enciclopedie, manuali, guide di vario tipo e libri informativi per diversi target. Per le mie analisi mi sono servito di 15 testi italiani e di 15 testi danesi.

Le dimensioni di questi due corpora sono citate nella Tabella 3.

Tabella 3 - *Corpus di Mr. Bean e SugarTexts: numero di parole e di proposizioni*

Corpus	Testi	1. Parole	2. Proposizioni	3. Periodi	4. Parole per periodo	5. Proposizioni per periodo	
						Num. medi	Differ.
<i>Mr. Bean</i>	<i>ital.</i>	7.278	1.388	319	22,8	4,35	32,6 %
	<i>dan.</i>	7.262	1.189	363	20,0	3,28	
<i>Sugar Texts</i>	<i>ital.</i>	4.819	898	194	24,8	4,63	82,3 %
	<i>dan.</i>	4.851	832	327	14,8	2,54	

Anche qui, i testi italiani e quelli danesi sono, in entrambi i corpora, grosso modo di dimensioni equivalenti quanto al numero di parole e di proposizioni (colonne 1-2). Invece i periodi italiani sono più lunghi quanto al numero di parole per periodo (colonna 4), e soprattutto quanto al numero di proposizioni per periodo (colonna 5), le differenze essendo particolarmente evidenti nei testi espositivi.

### 5. La testualizzazione delle singole proposizioni

Non vi è dubbio che un alto numero di proposizioni nello stesso periodo aumenti la complessità e la compattezza del periodo, ma un al-

tro fattore importante è la testualizzazione delle stesse proposizioni. Fra gli elementi che accrescono la complessità delle “unità di pensiero” (cioè dei periodi nella lingua scritta), Chafe (1985: 108-117) elenca frasi subordinate e apposizioni, participi presenti e passati e nominalizzazioni. Di tali testualizzazioni, i verbi impliciti e nominalizzati – ossia le testualizzazioni “deverbalizzate” (Korzen 2009, 2018; Korzen & Gylling 2017) – sono particolarmente interessanti perché aumentano anche la cosiddetta “densità testuale” e rendono i testi meno “vincolanti” nella terminologia di Sabatini (1999).

La densità testuale è definita come la relazione tra la quantità di materiale linguistico di una sequenza testuale e le informazioni che tale sequenza intende esprimere (Fabricius-Hansen 1996, 1998, 1999; Jansen 2003; Hansen-Schirra *et al.* 2007; Korzen & Gylling 2017), e le forme verbali non finite trasmettono un contenuto testuale con meno materiale linguistico rispetto ad una struttura con verbo finito, la quale in genere contiene una congiunzione e possibilmente un verbo ausiliario finito e un soggetto (Korzen 2014, 2015):

- (1) a. *Arrivato tardi*, Luca ha perso l’inizio del film.  
b. *Dato che Luca era arrivato tardi*, ha perso l’inizio del film.
- (2) a. *Arrivando tardi*, perderai l’inizio del film.  
b. *Se tu arrivi tardi*, perderai l’inizio del film.

Allo stesso tempo, le strutture non finite sono meno “vincolanti” perché lasciano un’interpretazione semantica precisa al ricevente (Sabatini 1999), ad esempio causa o condizione come illustrato negli esempi (1a)-(2a). In (3a) la nominalizzazione esprime un contenuto temporale:

- (3) a. *All’arrivo di Luca*, siamo andati al cinema.  
b. *Quando Luca è arrivato*, siamo andati al cinema.

Per queste ragioni tali strutture sono fortemente sconsigliate per esempio nei manuali e guide alla redazione dei documenti amministrativi e legislativi di tutti i livelli: comunale, provinciale, regionale, statale ed europeo (Korzen 2015). La maggiore densità linguistica e i meno vincoli interpretativi richiedono un maggiore impegno interpretativo da parte del ricevente (Korzen 2021), insomma comportano una maggiore complessità<sup>5</sup>.

<sup>5</sup> Cfr. anche Merlini Barbaresi (2003: 40ff, 2004: sezione 6) per il ruolo delle forme verbali non finite per la complessità testuale nelle ricette di cucina.

In un paragone cross-linguistico basato sui tre corpora indagati le testualizzazioni implicite e nominalizzate puntano su differenze notevoli:

Tabella 4 - *Corpus Europarl, Mr. Bean, SugarTexts: testualizzazione delle proposizioni*

<i>Corpus</i>	<i>Testi</i>	1. <i>Frase principali</i>	2. <i>Frase subordinate esplicite</i>	3. <i>Frase subordinate implicite</i>	4. <i>Nominalizzazioni</i>
<i>Europarl</i>	<i>italiani</i>	21,3 %	37,6 %	22,5 %	18,6 %
	<i>danesi</i>	28,1 %	49,2 %	12,4 %	10,3 %
<i>Mr. Bean</i>	<i>italiani</i>	40,9 %	19,5 %	33,7 %	5,9 %
	<i>danesi</i>	54,6 %	27,2 %	15,7 %	2,4 %
<i>SugarTexts</i>	<i>italiani</i>	29,8 %	17,8 %	26,7 %	25,6 %
	<i>danesi</i>	50,1 %	29,4 %	8,9 %	11,5 %

Nelle colonne 3-4 si osserva come, in tutti e tre i corpora, le testualizzazioni implicite e nominalizzate siano assai più frequenti nei testi italiani che nei testi danesi. Invece in tutti i testi danesi le percentuali di testualizzazioni esplicite, principali o subordinate (colonne 1-2), superano quelle italiane. Buoni esempi di testualizzazioni dense e – almeno per un danese – complesse perché consistono di periodi lunghi, compatti e pieni di verbi impliciti e nominalizzati, sono citati in (4)-(5) che provengono da due testi italiani del *corpus Europarl*.

- (4) Accanto al deciso sostegno ad un approccio microeconomico, destinato ad incoraggiare i paesi più poveri ad investire nel loro stesso avvenire lo sviluppo del microcredito, l'Unione auspica il mantenimento delle preferenze commerciali con i paesi più poveri e più economicamente vulnerabili. (ep-98-04-01.txt:39, deputato autore: Amadeo Amadeo)<sup>6</sup>.
- (5) [Q]uanto più noi riconosciamo l'autorevolezza degli organi monetari in questione – ed io mi associo a questo riconoscimento, non senza rilevare con convinzione l'esigenza di un governo politico dell'economia, anche a livello europeo – tanto più occorre pretendere che vengano dati esempi limpidi di efficienza, di moderazione salariale e di trasparenza. (ep-01-01-16.txt:39, deputato autore: Gianni Pittella).

<sup>6</sup> Le sigle riferenti ai discorsi *Europarl* ("ep") indicano anno-mese-giorno, seguiti dal numero del discorso del giorno in questione ("txt").



Invece gli esempi danesi in (6)-(7), dello stesso corpus, illustrano testualizzazioni completamente diverse, molto più verbali e semplici: composte di periodi brevi con frasi esplicite e un'unica subordinazione del primo livello.

(6) Der er mange lobbyister i Europa-Parlamentet. De indgår som en naturlig del af vores arbejde og bidrager med oplysninger og synspunkter. De kan nok ikke undværes, men vi skal have regler for deres aktiviteter. Ford-betænkningen er et godt bud på nogle regler, der kan gennemføres. (ep-96-07-16.txt:258, deputato autore: Freddy Blak).

‘Ci sono molti lobbisti al Parlamento europeo. Costituiscono una parte naturale del nostro lavoro e contribuiscono con informazioni e opinioni. Probabilmente non se ne può fare a meno, ma dobbiamo avere regole per le loro attività. La relazione Ford è una buona proposta per alcune regole che possono essere implementate.’

(7) I øvrigt løser direktivforslagene ikke forbrugernes problemer med handel over grænserne. Der er store afstande mellem forbrugere og sælgere i forskellige lande, både geografisk og sprogligt. Derfor kan man spørge sig selv, om harmonisering overhovedet har noget formål og nogen virkning på dette område. Jeg ønsker nemlig ikke harmonisering for harmoniseringens skyld. (ep-98-03-09.txt:60, deputata autrice: Ulla Sandbæk).

‘Inoltre le direttive proposte non risolvono i problemi dei consumatori con il commercio attraverso le frontiere. Ci sono grandi distanze tra consumatori e venditori in paesi diversi, sia geograficamente che linguisticamente. Perciò ci si può chiedere se l'armonizzazione abbia affatto uno scopo e un effetto in questo settore. Non desidero l'armonizzazione per il bene dell'armonizzazione.’<sup>7</sup>

---

<sup>7</sup> Queste due traduzioni di (6)-(7) sono mie e molto fedeli ai testi fonte. Le traduzioni ufficiali del Parlamento Europeo, scaricabili dall'Europarl Parallel Corpus 1996-2011, sono le seguenti:

(6') ‘Ci sono molti lobbisti al Parlamento Europeo. Costituiscono ormai un elemento naturale del nostro lavoro al quale contribuiscono con informazioni e punti di vista. Non se ne può fare a meno, ma occorre avere regole che disciplinino le loro attività. La relazione Ford rappresenta un ottimo spunto per regole possibili.’

(7') ‘Inoltre la proposta di direttiva non risolve il problema del consumatore rispetto agli acquisti effettuati all'estero. La distanza tra venditore e consumatore di paesi diversi è enorme, sia geograficamente che linguisticamente. Pertanto è lecito chiedersi

## 6. *Conclusioni*

La complessità linguistica è un fenomeno estremamente intricato, e come molti studiosi hanno affermato, probabilmente è impossibile determinare e definire una nozione quale la “complessità complessiva” di una lingua; nella migliore delle ipotesi, essa può essere considerata come l’accumulo di una serie di valori diversi. Nelle pagine precedenti ho messo a fuoco due tali valori, cioè il numero di proposizioni per periodo e il grado della loro deverbalizzazione, due fenomeni che si dimostrano particolarmente differenti nelle due lingue qui indagate, l’italiano e il danese, e particolarmente problematici per un danese che voglia perfezionare il suo italiano.

Anche se è possibile trovare testualizzazioni italiane e danesi strutturalmente diverse da quelle dimostrate rispettivamente in (4)-(5) e in (6)-(7), questi esempi costituiscono buone illustrazioni delle differenze cross-linguistiche citate nelle Tabelle 1-4 sopra, e gli esempi (4)-(5) confermano la particolare complessità strutturale che un apprendente danese deve affrontare nel suo incontro con l’italiano. Viceversa, la forma della testualizzazione danese vista in (6)-(7) può apparire banale e semplicistica ad un italiano.

Naturalmente l’indagine dei due fenomeni trattati in queste pagine non può per niente esaurire la problematica della complessità linguistica, ma non vi è dubbio che le citate differenze di compattezza e di densità testuale, documentate in modo palese dai corpora, siano una buona ragione per cui un danese possa trovare l’italiano “una lingua complessa”.

## *Riferimenti bibliografici*

Berruto, Gaetano. 2012. L’italiano popolare e la semplificazione linguistica. In Berruto, Gaetano, *Saggi di sociolinguistica e linguistica* a cura di Giuliano Bernini & Bruno Moretti & Stephan Schmid & Tullio Telmon, con la collaborazione di Gloria Scarano, 141–181. Alessandria: Edizioni dell’Orso. [Prima pubblicato in *Vox Romanica* 42 (1983). 38–79].

---

se l’armonizzazione in questo campo possa avere un senso e produrre degli effetti. Personalmente non sono a favore dell’armonizzazione come valore assoluto.’

Si noti come entrambe queste traduzioni siano più strutturalmente “eleganti” dei testi fonte e più sintatticamente complessa: in (6’) con una subordinazione in più e in (7’) con un verbo implicito (*effettuati*) in più.

- Bertucelli, Marcella. 2003. Cognitive complexity and the lexicon. In Merlini Barbaresi, Lavinia (a cura di), *Complexity in Language and Text*, 67–115. Pisa: Edizioni Plus.
- Chafe, Wallace L. 1985. Linguistic differences produced by differences between speaking and writing. In Olson, David R. & Torrance, Nancy & Hildyard, Angela (eds.), *Literacy, Language and Learning. The Nature and Consequences of Reading and Writing*, 105–123. Cambridge: Cambridge University Press.
- Dahl, Östen. 2009. Testing the assumption of complexity invariance: the case of Elfdalian and Swedish. In Sampson, Geoffrey & Gil, David & Trudgill, Peter (eds.), *Language Complexity as an Evolving Variable*, 50–63. Oxford: Oxford University Press.
- Deutscher, Guy. 2009. “Overall complexity”: a wild goose chase? In Sampson, Geoffrey & Gil, David & Trudgill, Peter (eds.), *Language Complexity as an Evolving Variable*, 243–251. Oxford: Oxford University Press.
- Ellis, Nick C. 2016. Salience, cognition, language complexity, and complex adaptive systems. *Studies in Second Language Acquisition* 38. 341–351. doi:10.1017/S027226311600005X
- Fabricius-Hansen, Cathrine. 1996. Informational density – a problem for translation and translation theory. *Linguistics* 34. 521–565.
- Fabricius-Hansen, Cathrine. 1998. Information density and translation, with special reference to German – Norwegian – English. In Johansson, Stig & Oksefjell, Signe (eds.), *Corpora and Cross-linguistic Research: Theory, Method, and Case Studies*, 197–234. Amsterdam: Rodopi.
- Fabricius-Hansen, Cathrine. 1999. Information packaging and translation. Aspects of translational sentence splitting (German – English/Norwegian). In Doherty, Monika (ed.), *Sprachspezifische Aspekte der Informationsverteilung*, 175–213. Berlin: Akademie-Verlag.
- Fiorentino, Giuliana. 2009. Complessità linguistica e variazione sintattica. *Studi Italiani di Linguistica Teorica e Applicata* XXXVIII(2). 281–312.
- Hansen-Schirra, Silvia & Neumann, Stella & Steiner, Erich. 2007. Cohesive explicitness and explicitation in an English-German translation corpus. *Languages in Contrast* 7(2). 241–265.
- Jansen, Hanne. 2003. *Densità informativa: Tre parametri linguistico-testuali. Uno studio contrastivo inter- ed intralinguistico*. Copenhagen: Museum Tusulanum Press.
- Koehn, Philipp. 2005. Europarl: A parallel corpus for statistical machine translation. *Conference Proceedings: The Tenth Machine Translation Summit*, 79–86. Thailand: Phuket.

- Korzen, Iørn. 2009. Struttura testuale e anafora evolutiva: tipologia romanza e tipologia germanica. In Korzen, Iørn & Lavinio, Cristina (a cura di), *Lingue, Culture e Testi Istituzionali*, 33–60. Firenze: Franco Cesati.
- Korzen, Iørn. 2014. Struttura testuale e anafora nella traduzione del discorso politico: un'indagine tipologico-comparativa. In Garavelli, Enrico & Suomela-Härmä, Elina (a cura di), *Dal manoscritto al web: canali e modalità di trasmissione dell'Italiano. Tecniche, materiali e usi nella storia della lingua: Atti del XII Congresso della Società Internazionale di Linguistica e Filologia Italiana (SILFI)*, 391–400. Firenze: Franco Cesati.
- Korzen, Iørn. 2015. Frasi complesse e complessità frasale: il discorso politico in un'ottica tipologico-comparativa. In Bruno, Carla & Casini, Simone & Gallina, Francesca & Raymond Siebetchu (a cura di), *Plurilinguismo/Sintassi. Atti del XLVI Congresso Internazionale della Società di Linguistica Italiana (SLI)*, 625–642. Roma: Bulzoni.
- Korzen, Iørn. 2018. L'italiano: una lingua esocentrica. Osservazioni lessicali e testuali in un'ottica tipologico-comparativa. In Korzen, Iørn (a cura di), *La Linguistica Italiana nei Paesi Nordici. Studi Italiani di Linguistica Teorica e Applicata*, XLVII(1). 15–36.
- Korzen, Iørn. 2021. Are some languages more complex than others? On text complexity and how to measure it. In Gargiulo, Marco & Haukås, Åsta & Korzen, Iørn (eds.), *When language typology meets multilingualism. From languages to uses and people. Globe. A Journal of Language, Culture and Communication* 12. 18–31.
- Korzen, Iørn & Gylling, Morten. 2017. Text structure in a contrastive and translational perspective: On information density and clause linkage in Italian and Danish. In Czulo, Oliver & Hansen-Schirra, Silvia (eds.), *Crossroads between Contrastive Linguistics, Translation Studies and Machine Translation*, 31–64. Berlin: Language Science Press. <https://zenodo.org/record/1019687>.
- Maas, Utz. 2009. Orality versus literacy as a dimension of complexity. In Sampson, Geoffrey & Gil, David & Trudgill, Peter (eds.), *Language Complexity as an Evolving Variable*, 164–177. Oxford: Oxford University Press.
- Masi, Silvia. 2003. The literature on complexity. In Merlini Barbaresi, Lavinia (a cura di), *Complexity in Language and Text*, 117–145. Pisa: Edizioni Plus.
- McWhorter, John H. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5. 125–166.

- Merlini Barbaresi, Lavinia. 2003. Towards a theory of text complexity. In Merlini Barbaresi, Lavinia (a cura di), *Complexity in Language and Text*, 23–66. Pisa: Edizioni Plus.
- Merlini Barbaresi, Lavinia. 2004. Levels of text complexity. In van Sterkenburg, Piet (ed.), *Linguistics Today – Facing a Greater Challenge*. Amsterdam/Philadelphia: John Benjamins, CD-Rom.
- Merlini Barbaresi, Lavinia. 2005. Il discorso economico/argomentativo: marcatezza e complessità della previsione. In Leandro Schena, Chiara Preite & Vecchiato, Sara (a cura di), *Gli insegnamenti linguistici nel nuovo ordinamento: Lauree triennali e specialistiche dell'area economico-giuridica*, 301–324. Milano: Egea.
- Miestamo, Matti. 2009. Implicational hierarchies and grammatical complexity. In Sampson, Geoffrey & Gil, David & Trudgill, Peter (eds.), *Language Complexity as an Evolving Variable*, 80–97. Oxford: Oxford University Press.
- Moretti, Bruno. 2018. Che cosa ha da dire la sociolinguistica sul tema della complessità delle lingue. *Rivista Italiana di Dialettologia* 42. 35–52.
- Nichols, Johanna. 2009. Linguistic complexity: a comprehensive definition and survey. In Sampson, Geoffrey & Gil, David & Trudgill, Peter (eds.), *Language Complexity as an Evolving Variable*, 110–125. Oxford: Oxford University Press.
- Sabatini, Francesco. 1999. “Rigidità-esplicitezza” vs “elasticità-implicitezza”: possibili parametri massimi per una tipologia dei testi. In Skytte, Gunver & Sabatini, Francesco (a cura di), *Linguistica testuale comparativa*, 141–172. Copenhagen: Museum Tusulanum Press.
- Skytte, Gunver & Korzen, Iørn & Polito Paola & Strudsholm, Erling (a cura di). 1999. *Tekststrukturering på italiensk og dansk. Resultater af en komparativ undersøgelse / Strutturezione testuale in italiano e danese. Risultati di una indagine comparativa*. Copenhagen: Museum Tusulanum Press.
- Smith, Viktor. 2009. Telling the SugarStory in seven Indo-European languages. What may and what must be conveyed? In Korzen, Iørn & Lavinio, Cristina (a cura di), *Lingue, culture e testi istituzionali*, 61–76. Firenze: Franco Cesati.