

La complessità morfologica Ricerca e didattica

Gabriele Pallotti

1. La complessità linguistica

I termini complesso e complessità sono usati frequentemente in relazione alla descrizione e all'insegnamento delle lingue, sia nelle discussioni scientifiche che nella conversazione ordinaria. Questa alta frequenza d'uso dovrebbe come minimo mettere in guardia sulla loro utilizzabilità come costrutti scientifici, i quali, per essere impiegati rigorosamente, devono avere un significato e un ambito di applicazione ben definiti.

Anche a livello dizionariale, il termine 'complesso' ha due sensi fondamentali:

1. Che presenta difficoltà per la comprensione o l'orientamento, dovute a profondità od oscurità di concetti oppure a una molteplicità di elementi o di aspetti [...] 2. Risultante di due o più parti interdipendenti (contrario di *semplice*).¹

Il primo senso implica difficoltà, sforzo cognitivo per un soggetto umano; il secondo senso descrive invece la complessità intrinseca, strutturale, di un oggetto. Proporrò di chiamare il primo concetto difficoltà, mentre si può riservare il termine complessità al secondo, che riguarda le caratteristiche strutturali. Certamente, la 'molteplicità di elementi o di aspetti' (complessità strutturale) può essere responsabile della 'difficoltà per la comprensione o l'orientamento' (complessità cognitiva), e questo legame può diventare l'oggetto di ricerche empiriche, ma piuttosto che usare lo stesso termine per designare la causa e l'effetto, è meglio utilizzare due termini distinti.

In questo saggio si proporrà un modo esplicito di calcolare la complessità morfologica nei testi scritti, detto Indice di Complessità Morfologica (ICM). Esso verrà utilizzato dapprima per confrontare la complessità morfologica in tre lingue diverse, e in seguito per l'analisi delle produzioni di scrittori nativi e non-nativi dell'italiano. Infine, si discuterà come sia possibile utilizzare l'ICM nella didattica dell'italiano come lingua seconda/straniera.

Negli ultimi decenni si è assistito a un vivace dibattito nella linguistica teorica e tipologica sulla nozione di complessità in diversi sistemi linguistici. Tale dibattito rimette in discussione quello che alcuni² hanno definito un assioma della linguistica del XX secolo, cioè che tutte le lingue sono ugualmente complesse. Esso era a sua volta una risposta a un altro pregiudizio del secolo precedente, e cioè che le lingue non sono solo diverse tra loro, ma sono anche gerarchicamente ordinate, dalle più primitive alle più evolute, dalle più semplici alle più sofisticate, venendo queste ultime identificate da molti con le lingue europee con ricchi paradigmi morfologici, vive come il tedesco

¹ Cfr. Devoto-Oli 2000, s. v. *complesso*.

² Cfr. Sampson 2009: 1-18.

o il francese, o morte come il latino e il greco.³ A dire il vero, già nel Settecento e Ottocento alcuni, tra cui Herder e Humboldt, avevano messo in discussione queste graduatorie ingenuie, mostrando come anche le lingue di popoli cosiddetti primitivi avessero una grande ricchezza di categorie, che consentivano a loro volta di esprimere le più varie sfumature del pensiero, in modi spesso inaccessibili alle lingue europee. Queste idee sono state in seguito sviluppate in un programma rigoroso di ricerca scientifica da Boas, uno dei fondatori della linguistica moderna, e poi ulteriormente divulgate da Sapir e Whorf.⁴

A metà del Novecento, dunque, il dibattito sulla superiorità o inferiorità delle lingue era considerato definitivamente chiuso, un retaggio di una para-linguistica prescientifica da cui bisognava prendere le distanze nel modo più netto possibile. Le discussioni sulla maggiore o minore complessità dei sistemi linguistici parevano essere collegate a tale retaggio, ed erano perciò sbrigativamente liquidate con la professione di fede nel fatto che tutte le lingue sono ugualmente complesse: se una ha una morfologia particolarmente ricca e intricata, sarà semplice nella fonologia o nella sintassi, e viceversa, di modo che alla fine la complessità totale del sistema rimanga sempre costante.

Intorno alla fine del XX secolo, però, il dibattito si è riaperto: non certo nei termini della superiorità o inferiorità dei sistemi linguistici, ma in quelli più neutri e oggettivi della loro complessità. Trudgill 1989,⁵ ad esempio, è stato tra i primi a rilevare che le lingue che nella loro storia sono state apprese da un gran numero di parlanti non nativi, come l'inglese, tendono ad essere più semplici di quelle parlate in piccole comunità isolate. McWhorter 2001⁶ ha esplicitato ulteriormente questa affermazione, affermando nel titolo di un suo articolo che 'The world's simplest grammars are creole grammars'. Questo punto di vista è stato ulteriormente sviluppato nel volume *Language interrupted*,⁷ in cui si discutono diverse lingue, tra cui i creoli come caso estremo, la cui trasmissione intergenerazionale è stata interrotta da un gran numero di apprendenti non nativi, il che ha portato alla nascita di sistemi linguistici 'più semplici'. Negli anni successivi c'è stata un'esplosione di pubblicazioni sul tema, con decine di articoli e diversi volumi (ad es. Miestamo *et al.*, 2008; Sampson *et al.*, 2009; Kortmann & Szendrői, 2012), che non è possibile riassumere in questa sede. Basti dire che la maggior parte dei lavori riguarda la complessità dei sistemi linguistici, ad esempio dei paradigmi morfologici o dei repertori fonologici, e solo una piccola proporzione ha invece a che fare con la misurazione della complessità così come essa si manifesta nei testi concreti, che è invece l'approccio generalmente seguito nella linguistica acquisizionale e in questo saggio in particolare, che confronterà diversi testi prodotti in tre diverse lingue europee.

2. La complessità nell'apprendimento delle lingue

Anche negli studi sull'apprendimento e insegnamento delle lingue la complessità è un tema importante. Essa fa parte di una triade di misure che vengono comunemente usate per descrivere diverse dimensioni dello sviluppo linguistico: Complessità, Accuratezza, Fluency (o CAF). Già negli anni Settanta, con l'avvento degli approcci comunicativi, si riconosceva che l'evoluzione della seconda lingua non può essere caratterizzata solo in termini di accuratezza, come decremento del numero di errori, ma che questa dimensione doveva essere perlomeno integrata da quella della fluency.⁸

³ Cfr. Eco 1992.

⁴ Per una panoramica storica su questa linea di pensiero, cfr. Pallotti 1999: 109-138.

⁵ Trudgill 1989: 227-237.

⁶ McWhorter 2001: 125-166.

⁷ Cfr. McWhorter 2007.

⁸ Cfr. Brumfit & Johnson 1979.

Un paio di decenni più tardi, Skehan⁹ proponeva di aggiungere una terza dimensione chiave, la complessità: a volte i sistemi interlinguistici crescono non tanto perché diminuisce il numero di errori (accuratezza) o aumenta la fluenza, ma perché essi vengono ristrutturati in modo da divenire più complessi, con nuove strutture e nuove regole.

Termini come complesso e complessità figurano di frequente nel discorso degli insegnanti di lingue, che li usano per riferirsi tanto alla varietà delle strutture conosciute e prodotte dai propri allievi, quanto alla loro difficoltà e sofisticatezza.¹⁰ I termini appaiono anche ripetutamente nel Quadro Comune Europeo di Riferimento per le lingue (QCER):¹¹ nell'edizione inglese, Bernardini e Granfeldt 2014¹² hanno conteggiato 101 occorrenze di *complex* e 13 di *complexity*, accanto a 259 di *simple* e 70 di *basic*. La parola *complex* è la 155ma per frequenza nell'intero testo, e la 120ma nel capitolo 5, dove vengono descritte le competenze degli apprendenti. Si tratta dunque di una nozione che pare avere un peso considerevole nel discorso meta-didattico.

Anche nel campo della ricerca sull'acquisizione della L2 gli studi sono stati numerosi e hanno esplorato diverse direzioni, quali lo sviluppo di complessità, accuratezza e fluenza nel tempo, o il loro rapporto con le caratteristiche delle situazioni comunicative (*task*).¹³ Per quanto riguarda specificamente la complessità, è stata sottolineata più volte la polisemia e 'complessità' del costrutto stesso: esso può riferirsi tanto a caratteristiche della situazione comunicativa quanto alle produzioni linguistiche stesse; inoltre, come si è già notato, può significare tanto intricatezza strutturale quanto difficoltà cognitiva.¹⁴ Infine, sono state proposte decine di misure diverse della complessità negli studi sulla L2, spesso con scarsa attenzione al fatto che alcune di esse si sovrappongono, altre si contraddicono a vicenda, altre ancora misurano diversi aspetti del costrutto, e risulta perciò difficile alla fine comparare gli studi e trarre conclusioni di carattere generale.¹⁵ In linea di massima, le ricerche sembrano indicare che, per quanto riguarda la complessità lessicale,¹⁶ essa tenda ad aumentare con lo sviluppo della L2. Questo implica infatti un ampliamento del lessico e dunque un maggior numero di tipi lessicali diversi rispetto alle occorrenze prodotte.¹⁷

Per quanto riguarda la complessità sintattica, i risultati sono un po' meno lineari. Pare abbastanza assodato che, ai livelli iniziali, le strutture sintattiche tendano a essere semplici, cioè brevi, paratattiche, scarsamente integrate tra loro. In seguito aumentano misure come la lunghezza media della clausola o della T-Unit,¹⁸ che indicano progetti sintattici di maggiore portata. Ad esempio, Gyllstad, Granfeldt, Bernardini e Källkvist 2014¹⁹ hanno mostrato che la complessità sintattica (misurata come lunghezza media in parole delle T-Unit e come numero di unità sintattiche per T-Unit) tenda ad aumentare nelle produzioni di apprendenti non nativi dell'inglese, francese e italiano, a seconda che essi si trovino ai livelli A o B del QCER. D'altra parte, altre ricerche mostrano che questo aumento della complessità sintattica non è

⁹ Cfr. Skehan 1998. Si veda anche Wolfe-Quintero, Inagaki & Kim, 1998.

¹⁰ Ma, a livello di ricerca scientifica, si raccomanda perlopiù di utilizzare per queste nozioni termini distinti, quali complessità e difficoltà: cfr. Pallotti 2015a: 117-134, Bulté & Housen 2012: 21-46.

¹¹ Cfr. Consiglio d'Europa 2001.

¹² Bernardini & Granfeldt 2014: 10-15.

¹³ Per rassegne, Housen, Kuiken & Vedder 2012; Long 2015; Michel 2017: 50-68.

¹⁴ Pallotti 2009: 590-601; Bulté & Housen 2012: 21-46.

¹⁵ Norris & Ortega 2009: 555-578; Bulté & Housen 2012: 21-46; Pallotti 2015a: 117-134.

¹⁶ La complessità lessicale, in molte ricerche acquisizionali e in un approccio puramente strutturale come il nostro, si può intendere come varietà dei lessemi usati in un testo; cfr. Pallotti 2015.

¹⁷ Cfr. De Clercq 2015: 69-94.

¹⁸ Frase principale con tutte le sue dipendenti, cfr. Hunt 1965.

¹⁹ Cfr. Gyllstad, Granfeldt, Bernardini & Källkvist 2014: 1-30.

continuo fino ai livelli più avanzati. Ad esempio, apprendendo varietà di prosa accademica più sofisticate, che contengono molte nominalizzazioni e sintagmi complessi, queste strutture tendono ad aumentare, riducendo però la complessità sintattica a livello di subordinazione.²⁰ Inoltre, almeno nelle interazioni orali, una maggiore fluency interazionale porta a produrre turni più brevi, sintatticamente semplici, nelle varietà più avanzate, mentre quelle intermedie sono caratterizzate da una sintassi più complessa che corrisponde a dinamiche interazionali meno fluide.²¹

Molto più limitate sono state le ricerche sulla complessità morfologica. Fino ad alcuni anni fa, i pochi studi che hanno provato a misurarla si sono basati su indicatori piuttosto grossolani, come la frequenza delle forme verbali flesse, il numero di diverse forme verbali o la varietà delle forme verbali passate.²² Oltre a essere parziali, tutti questi indici hanno il problema di esprimere valori assoluti (quantità, frequenze), che, come è noto, sono estremamente sensibili alla lunghezza del testo. Più recentemente, Pallotti²³ ha proposto una definizione operativa del costrutto che risente meno di questi inconvenienti, e che sarà presentata, con alcune modifiche, nelle prossime pagine, insieme a uno dei primi studi in cui è stata applicata all'acquisizione dell'italiano L2. Un altro lavoro che ha misurato la complessità morfologica nell'apprendimento della L2 è quello di De Clercq e Housen 2017,²⁴ basato su racconti narrativi orali prodotti da parlanti fiamminghi che apprendevano il francese e l'inglese come seconde lingue. Gli autori hanno osservato lo sviluppo della complessità dei paradigmi verbali, usando tre diversi indici di complessità: l'Indice di Complessità Morfologica,²⁵ l'Inflectional Diversity index (ID)²⁶ e l'indice Types/Family.²⁷ Con tutte e tre le misure, risulta che nel francese L2 la complessità morfologica tende a crescere in misura maggiore, fino ai livelli più avanzati, rispetto a quanto accade con l'inglese L2. La prima misura sarà presentata in dettaglio in questo articolo; la seconda consiste nel sottrarre la diversità dei lemmi (varietà lessicale) dalla diversità delle forme flesse, per fornire un'indicazione indiretta della diversità delle forme flessive; la terza riguarda la morfologia derivazionale e calcola il numero di forme derivate da una parola-base. In particolare, la complessità morfologica degli apprendenti dell'inglese aumenta soprattutto a causa dell'introduzione di un maggior numero di verbi irregolari, mentre quella del francese cresce sia per gli irregolari che per l'introduzione di un maggior ventaglio di forme dai paradigmi dei verbi regolari. In entrambe le lingue, la crescita appare più pronunciata nelle fasi iniziali-intermedie di apprendimento, e meno in quelle più avanzate, anche se non si arresta mai completamente.

3. Misurare la complessità morfologica: l'ICM

Per misurare la complessità morfologica in modo oggettivo e replicabile su lingue diverse, è stato sviluppato l'Indice di Complessità Morfologica,²⁸ per ora applicato esclusivamente alla morfologia verbale, anche se sarebbe possibile, in linea di principio, estenderlo a quella nominale e di altre classi di parole, purché presentino una variazione di tipo formale causata da processi morfologici. La scelta di concentrarsi per ora sulla morfologia verbale è dettata essenzialmente da ragioni di praticità e limitatezza di risorse; dato che, nelle lingue europee esaminate fino ad ora,

²⁰ Cfr. Norris & Ortega 2009: 555-578.

²¹ Cfr. Pallotti & Ferrari 2008; Nuzzo & Gauci 2012.

²² Per una rassegna, cfr. Bulté & Housen 2012: 21-46.

²³ Pallotti 2015a: 117-134, 2015b: 195-218.

²⁴ De Clercq & Housen 2017: 315-334.

²⁵ Cfr. Pallotti 2015a: 117-134.

²⁶ Cfr. Malvern, Richards, Chipere & Durán 2004.

²⁷ Cfr. Horst & Collins 2006: 83-106.

²⁸ Cfr. ICM, Pallotti 2015a: 117-134, 2015b: 195-218.

la morfologia verbale è molto più ricca di quella nominale, si è preferito iniziare con questa. L'ICM si basa sull'idea che un testo in cui compaia una vasta gamma di forme verbali diverse tra loro sia più complesso, da questo punto di vista, di uno in cui si ripetano sempre le stesse forme. L'idea di fondo è la stessa che sta alla base degli indici di complessità lessicale basati sulla diversità dei lessemi. Da questo punto di vista, la complessità può essere definita come alta diversità dei tipi (*types*) con scarsa ripetizione delle occorrenze (*token*), cioè con un alto rapporto *type/token*. Dunque, un testo che contenga i tre diversi lessemi *leggere*, *scrivere*, *pensare* sarà più complesso sul piano lessicale di uno che contenga tre ripetizioni dello stesso lessema, come *pensare*, *pensare*, *pensare*, ma anche *pensare*, *penso*, *pensavate*.

La complessità morfologica (limitandoci alla morfologia flessiva) viene misurata, analogamente, calcolando il rapporto tra tipi di morfemi flessivi e le loro occorrenze, o ripetizioni. Dunque, un testo sarà più complesso sul piano morfologico se conterrà tre forme flesse come *pensare*, *penso*, *pensavate*, rispetto a uno che contenga tre volte la stessa forma flessa, come in *pensare*, *pensare*, *pensare*, ma anche in *pensare*, *parlare*, *mangiare*.

Per calcolare questa diversità, sia a livello lessicale che morfologico, si calcola il rapporto tra *type* e *token*, ovvero il Type-Token Ratio (TTR). Gli studi sulla diversità lessicale, tuttavia, hanno ripetutamente mostrato che questa misura è fortemente sensibile alla lunghezza del testo: nei testi più lunghi, è inevitabile che i lessemi tendano a ripetersi, per cui i valori del TTR tendono sistematicamente a calare all'aumentare della lunghezza del testo. Sono state proposte diverse soluzioni a questo problema, alcune anche molto complesse da un punto di vista tecnico. Una delle più semplici, che comunque rimane sempre valida, consiste nel calcolare la diversità media all'interno di sotto-campioni testuali di lunghezza costante, tipicamente 50 o 100 parole: si tratta dello Standardized Type/Token Ratio.²⁹

L'ICM si basa sullo stesso principio: si estraggono dal testo campioni di N forme verbali (10, negli studi svolti fino ad ora; il numero è stato scelto per la sua semplicità e perché consente di calcolare l'ICM su campioni contenenti un minimo di 20 verbi)³⁰ e si calcola la diversità media al loro interno. La procedura è in realtà leggermente più complicata: ne daremo ora una breve spiegazione, rimandando ad altri lavori per una trattazione più dettagliata.³¹ L'approccio generale seguito per calcolare l'ICM si basa su due passi consecutivi. Il primo passo consiste in un'analisi linguistica che individua all'interno di un testo gli esponenti morfologici, ovvero le diverse forme che possono assumere i lessemi appartenenti a una certa classe di parole. Più tecnicamente, sarebbe opportuno parlare di processi morfologici che trasformano una base lessicale in una forma flessa; per brevità, useremo il termine 'esponenti' per riferirci a questi processi morfologici.³² Il secondo passo consiste in un'analisi matematica che calcola la varietà degli esponenti all'interno dei campioni e tra di essi.

Per quanto riguarda l'analisi linguistica, è importante chiarire che l'ICM si basa sul computo degli esponenti, cioè delle forme flesse che possono assumere le parole. Da questo punto di vista, in *tu cant-i* (2sg.pres.indic) e in *che lui cant-i* (3sg.pres.cong) è presente un solo esponente, *-i*, anche se esso esprime sincreticamente diverse funzioni grammaticali (la seconda o terza persona; il modo indicativo o congiuntivo). D'altro canto, gli allomorfi che esprimono la stessa funzione grammaticale sono contati separatamente, per cui in *arriv-ato*, *pr-eso* e *rim-asto* sono presenti tre diversi esponenti, cioè diversi processi morfologici che modificano una forma lessicale di base

²⁹ Cfr. STTR, Johnson 1944: 1-15.

³⁰ Per un primo tentativo di validazione con campioni di diversa grandezza, si veda Pallotti 2015b: 195-218.

³¹ Oltre al sito del progetto, si vedano Pallotti 2015a: 117-134, 2015b: 195-218; Brezina & Pallotti 2016.

³² Si veda Haspelmath & Sims 2010 per una trattazione più approfondita.

per darle diverse realizzazioni flessive. In altre parole, si calcola il numero delle forme, non delle combinazioni forma/funzione, o celle del paradigma.

Questa scelta si basa, tra le altre cose, sul fatto che, nei testi prodotti da apprendenti di prima e seconda lingua, non è sempre possibile identificare l'esatta funzione di una forma flessa. Ad esempio, in *Antonio lascia lavoro* come si deve interpretare *lascia*? Come un presente indicativo, durativo/abituale, di terza persona singolare, cioè con il valore che ha in italiano, o come un presente generico, che copre anche significati progressivi, o persino di passato, in certi contesti, e che può applicarsi a diverse persone, forse a tutte quelle del paradigma?

Insomma, nella varietà di apprendimento non è sempre possibile stabilire nettamente le funzioni grammaticali, mentre le forme linguistiche sono in linea di massima più identificabili, ed è per questo che l'ICM, che si prevede possa essere usato frequentemente nelle ricerche acquisizionali, si limita a calcolare la varietà delle forme flessive. Se lo si desidera, in ogni caso, ma non ci risulta che sia stato ancora fatto, è possibile applicare la procedura matematica per analizzare la diversità dei rapporti forma-funzione, cioè delle celle dei paradigmi, calcolando, per ciascun campione, il numero di stringhe composte da esponente formale + significato grammaticale, quali 'i:2sg.pres.indic' o 'i:3sg.pres.cong'. Inoltre, sia che si calcoli la varietà delle forme o dei rapporti forma-funzione, l'approccio qui seguito si basa sull' "Interlanguage approach",³³ che esclude rigorosamente ogni confronto con la lingua target nell'interpretazione dell'interlingua, cioè ogni riferimento agli errori. In altri termini, un'interlingua può risultare più o meno complessa dal punto di vista dell'ICM senza essere necessariamente più o meno corretta dal punto di vista delle norme della lingua target: complessità e accuratezza sono costrutti diversi che potrebbero essere correlati empiricamente (ma non è detto: a volte si possono riscontrare aumenti della complessità con diminuzioni dell'accuratezza, o viceversa), ma che devono essere tenuti distinti concettualmente.

Anche per identificare gli esponenti, o i processi morfologici sottostanti a ciascuna forma flessa del verbo, si segue un approccio puramente formale. Per ogni verbo, si identifica una forma di base lessicale, la 'base di default',³⁴ che è quella che ricorre più volte nell'intero paradigma verbale della lingua target. Con i verbi regolari come *cantare* la sua identificazione è facile (*cant-*), mentre per quelli irregolari le cose si complicano. Ad esempio, la base di *prendere* è *prend-*, perché ricorre nella maggior parte delle forme del paradigma verbale, anche se alcune forme (minoritarie) si costruiscono sulla base *pres-* (*preso*, *presi*, *prese* ecc). Il processo morfologico che porta alla costruzione di una forma come *prendiamo* sarà descritto come la semplice aggiunta del suffisso *-iamo*; quello che produce *prese* sarà invece caratterizzato come un'operazione complessa sulla base che implica il cambiamento della sua quarta e quinta consonante (*nd*) con una nuova consonante (*s*) e l'ulteriore aggiunta del suffisso di tempo/persona *-ero*. Non si pretende naturalmente che queste operazioni di 'cambiamento della base' abbiano una validità storica, teorica o psicolinguistica: si tratta piuttosto di un algoritmo descrittivo (simile a ciò che in linguistica computazionale si chiama distanza di Levenshtein), che ha il pregio di essere oggettivo e applicabile a diverse lingue, indipendentemente dalle proposte teoriche su come analizzarne la struttura in radici, temi, morfemi.

Una volta identificati gli esponenti flessivi per ciascuna classe di parole all'interno di un testo, inizia l'analisi matematica della loro diversità. Dal numero totale degli esponenti si estraggono campioni di N forme, ad esempio dieci. Per ciascuno di questi campioni, si calcola il numero di diversi tipi flessivi, cioè di diverse

³³ Cfr. Pallotti 2017: 393-412.

³⁴ Cfr. Brezina & Pallotti 2016.

forme (min 1 - max 10), e si calcola poi la diversità media all'interno dei campioni. Viene poi calcolata la diversità tra campioni, identificando, per ogni coppia di campioni, gli esponenti unici, non condivisi (min 0 - max 20), e si trova di nuovo la media di questi valori, dividendola poi per due. Infine, il valore medio di diversità nei campioni viene sommato al valore medio di diversità tra campioni diviso due, si sottrae 1 e si arriva così all'Indice di Complessità Morfologica.

L'intera procedura per il calcolo dell'ICM può essere svolta attraverso un programma informatico sviluppato da Vaclav Brezina e Gabriele Pallotti, disponibile all'indirizzo http://corpora.lancs.ac.uk/vocab/analyse_morph.php, dove si trovano anche spiegazioni più dettagliate sulle modalità di identificazione degli esponenti e di calcolo della loro diversità. Il programma restituisce dapprima un'analisi morfologica del testo inserito, basandosi sul lemmatizzatore TreeTagger,³⁵ che consente di identificare, per ciascuna parola, la base, in modo che sia possibile estrarre dalla forma flessa l'esponente morfologico. La lista di esponenti così ottenuta in automatico può essere modificata liberamente dal ricercatore, per rettificare eventuali errori di codifica, dovuti il più delle volte a problemi nella lemmatizzazione automatica. È anche possibile importare una lista di esponenti estratti manualmente o con altre procedure.

Una volta costruita la lista di esponenti morfologici, il programma consente di eseguire la parte matematica dell'analisi, estraendo un numero di campioni casuali di N forme ciascuno e calcolando, per ciascuno di essi, il numero di diversi esponenti che contiene e il grado di diversità rispetto a un altro campione casuale. Questa procedura viene ripetuta un certo numero di volte,³⁶ per arrivare a stabilire la diversità media intra- e inter-campioni, su cui si basa il calcolo dell'ICM.

4. Studio 1: la complessità morfologica in diverse lingue

Una prima applicazione dell'ICM consiste nel misurare come la complessità morfologica (intesa come varietà degli esponenti verbali) vari tra diversi generi testuali nella stessa lingua o tra diverse lingue. Per fare ciò, sono stati analizzati sei diversi testi per cui erano disponibili traduzioni parallele in italiano, inglese e tedesco (le lingue sono state scelte per ragioni di praticità e in base alla conoscenza diretta di chi scrive; i testi perché rappresentano generi piuttosto diversi tra loro), in modo tale da controllare il più possibile l'omogeneità dei contenuti; anche la lunghezza è omogenea, essendo sempre intorno alle 1000 parole. Si tratta dei seguenti testi (tra parentesi, l'abbreviazione usata nel grafico):

- Quadro Comune Europeo di Riferimento per le lingue (QCER);³⁷ solo le scale dei descrittori generali, in prima e terza persona, cap. 3.1 (QCE)
- Risoluzione del Parlamento europeo del 14 marzo 2013 sulla tabella di marcia per l'energia 2050 (solo punti 1-10) (RIS)
- Articolo giornalistico pubblicato nella sezione notizie di attualità del Parlamento europeo: 'Il futuro della Grecia: il dibattito con il Primo ministro greco Alexis Tsipras', dell'08-07-2015 (ART)
- Un saggio politico di N. Chomsky (1992), 'What Uncle Sam really wants' (sez 1-2) (CHO)
- Favola di Cappuccetto Rosso (CAP)
- Vangelo di Marco, capp 1-2 (MAR)

³⁵ Cfr. Schmid 1994.

³⁶ Pallotti 2015b: 195-218 ha dimostrato che 100 rappresenta un numero sufficiente.

³⁷ Cfr. Consiglio d'Europa 2001.

	Inglese	Tedesco	Italiano
QCE	3.54	5.37	11.37
RIS	4.85	6.83	10.76
ART	7.11	9.54	11.47
CHO	7.15	11.16	15.91
CAP	8.94	13.73	13.63
MAR	9.16	14.02	15.8

Tabella 1 La complessità morfologica in diversi testi e diverse lingue

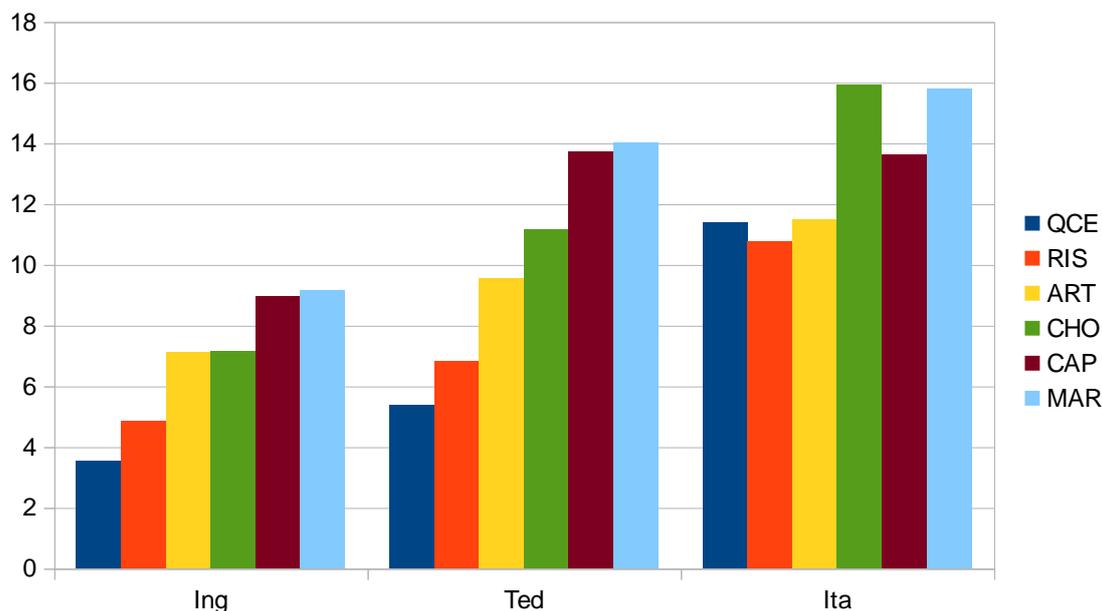


Figura 1 La complessità morfologica in diversi testi e diverse lingue

Come si vede nella Tabella 1 e in Figura 1, le tre lingue hanno chiaramente diversi livelli di complessità morfologica. Come prevedibile, la più bassa è quella dell'inglese, seguita dal tedesco e dall'italiano. Il livello più alto di complessità delle forme verbali in inglese, che viene raggiunto nell'estratto dal Vangelo di Marco (9.16), è in ogni caso inferiore al livello più basso riscontrato in italiano, cioè la Risoluzione del Parlamento europeo (10.76). È anche interessante notare come la complessità morfologica dei diversi testi segua un andamento identico in inglese e tedesco: in queste due lingue, esattamente gli stessi testi occupano le stesse posizioni nella gerarchia, dal più semplice al più complesso. Per quanto riguarda l'italiano, invece, si riscontrano lievi differenze: in particolare, le scale dei descrittori del QCER risultano essere leggermente più complesse della risoluzione del Parlamento europeo (è il contrario nelle altre due lingue), e i tre testi più complessi in italiano seguono l'ordine CAP, MAR, CHO, che è invece CHO, CAP, MAR nelle altre due lingue. Certo, c'è una sostanziale sovrapposizione, per cui si può dire che, in tutte e tre le lingue, alcuni testi, come Cappuccetto Rosso o il Vangelo risultano avere una complessità morfologica maggiore di altri testi relativamente semplici, come i descrittori del Quadro comune europeo o la risoluzione parlamentare. La ragione è da ricercarsi nel fatto che questi ultimi testi sono essenzialmente omogenei per quanto riguarda il tempo verbale (presente) e la persona (terza), mentre quelli più complessi – essendo narrazioni che contengono anche dialoghi – contengono una maggiore varietà di persone (la terza del narratore

ma anche la prima e la seconda dei dialoghi) e tempi e aspetti verbali (passato remoto e imperfetto per la trama e lo sfondo del racconto, presente e imperativo per i dialoghi).

Si può fare un'ultima osservazione su questi dati. Nella Figura 1 si vede in modo abbastanza evidente come la complessità dei diversi testi in inglese sia sempre piuttosto limitata, producendo una scala sostanzialmente schiacciata sui valori più bassi: la gamma di variazione va da 3.54 a 9.16, pari a 5.62 punti; in italiano, invece, è sempre abbastanza alta, con una gamma di variazione che va da un minimo di 10.76 a un massimo di 15.91, per un'estensione di 5.15 punti. Il tedesco è diverso da questo punto di vista: i suoi testi morfologicamente più semplici hanno valori molto bassi, paragonabili a quelli dell'inglese (minimo: 5.37), mentre quelli con maggiore complessità morfologica si avvicinano ai livelli dell'italiano (massimo: 14.02); la gamma di variazione risulta dunque essere più ampia, per un'estensione complessiva di 8.65 punti.

5. Studio 2: la complessità morfologica nell'apprendimento dell'italiano L2

Brezina e Pallotti 2016 hanno condotto uno studio che applica l'ICM ai testi prodotti da apprendenti dell'italiano e dell'inglese. In questa sede diamo una breve sintesi dei risultati relativi all'italiano, rimandando all'articolo originale per una trattazione più estesa.

Il corpus analizzato consiste di 18 testi prodotti da parlanti nativi e 39 testi prodotti da studenti universitari olandesi che seguono corsi di italiano come lingua straniera. I dati sono stati gentilmente forniti dai coordinatori del progetto 'Communicative Adequacy and Linguistic Complexity in L2 Writing' (CALC).³⁸ Ogni partecipante ha prodotto due brevi saggi argomentativi, che sono stati considerati un'unica produzione per gli scopi di questo lavoro, in modo da avere un numero sufficiente di forme verbali per calcolare in modo affidabile l'ICM: i due testi combinati producevano una lunghezza media di circa 250 parole. Il livello di competenza linguistica complessiva degli apprendenti è stato misurato mediante un C-test,³⁹ che richiedeva di completare una parola su due a cui era stata rimossa metà delle lettere (punteggio massimo = 100). In tal modo, è stato possibile suddividere gli apprendenti in due sottogruppi di uguale numerosità, contenenti rispettivamente quelli di livello più alto, con punteggio superiore al valore mediano di 71, e quelli con un punteggio più basso di tale soglia.

Corpus	Media (DS)	Gamma variazione
Nativi	12.85 (1.42)	10.60 - 15.33
Non nativi (tutti)	11.75 (1.93)	7 - 15.30
Non nativi - alto (> 71)	13.05 (1.15)	11.33 - 15.30
Non nativi - basso (≤ 71)	10.64 (1.78)	7 - 14.30

Tabella 2 ICM di parlanti nativi e non nativi dell'italiano

Come si vede nella Tabella 2, il gruppo degli apprendenti di livello più alto aveva dei valori di complessità morfologica del tutto paragonabili a quelli dei parlanti nativi; quelli con minori livelli di competenza, invece, producevano una varietà di forme verbali nettamente inferiore. È anche da notare che la varianza dei punteggi tra i nativi

³⁸ Cfr. Kuiken, Vedder & Gilabert 2010: 81-99; Kuiken & Vedder 2014: 329-348.

³⁹ Cfr. Klein-Braley & Raatz 1982: 23-37.

e tra gli apprendenti di livello più avanzato era relativamente limitata, mentre essa risultava considerevolmente più alta nel gruppo di apprendenti di livello più basso, i cui valori di ICM si estendevano su una gamma molto ampia (da 7 a 14.30) e con un'elevata deviazione standard (1.78 su una media di 10.64).

Guardando in modo più particolareggiato alla correlazione punto-per-punto tra valori dell'ICM e livelli di competenza linguistica misurati attraverso il C-test, si confermano questi risultati. La Figura 2 mostra graficamente la correlazione esistente tra punteggi del C-test e valori dell'ICM. Come si può notare, i due valori sono fortemente correlati, con un valore dell'indice di correlazione di Pearson molto alto ($r = 0.759$, $p < 0.001$, 95% CI [0.584, 0.867]).

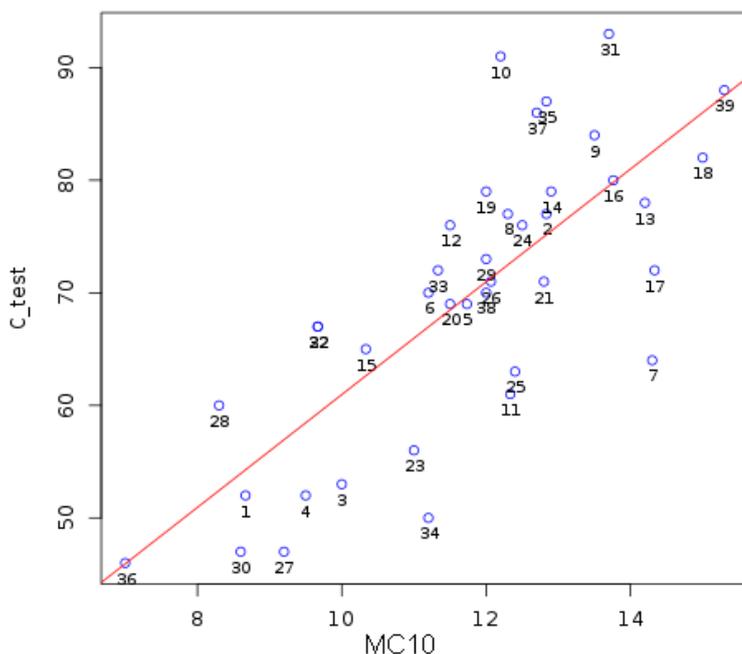


Figura 2 Correlazione tra valori dell'ICM (MC10) e punteggi nel C test - parlanti non nativi

La complessità morfologica risulta essere correlata anche ad altri indicatori di conoscenza della L2, anche se in modo meno netto. In particolare, si ha una correlazione positiva ($r = 0.441$, $p = 0.005$, 95% CI [0.145, 0.664]) con l'indice di diversità lessicale calcolato su campioni di uguale lunghezza (100 parole), ovvero lo Standardized Type/token ratio di cui si è parlato in precedenza, e con la complessità sintattica, misurata con la lunghezza media dei periodi da punto a punto ($r = 0.416$, $p = 0.008$, 95% CI [0.115, 0.646]).

Questi dati mostrano dunque che la complessità morfologica tende a crescere con l'aumentare delle competenze in italiano L2. In altri termini, apprendere la morfologia di una lingua con paradigmi verbali complessi quale è l'italiano richiede un considerevole investimento in termini di sviluppo della competenza linguistica, il quale va di pari passo con altri aspetti, quali la conoscenza di un lessico ampio e variato e la capacità di costruire strutture sintattiche più estese.

6. Implicazioni didattiche

Il costrutto di complessità morfologica e la sua definizione operativa sono stati sviluppati in primo luogo per finalità di ricerca scientifica, all'interno di un più ampio programma di indagine sui rapporti tra complessità linguistica, uso e apprendimento delle lingue e caratteristiche delle situazioni comunicative. Tuttavia, il costrutto e la sua misurazione pratica si prestano anche a essere utilizzati in contesti didattici: nelle prossime pagine proveremo a fornire alcuni esempi applicativi concreti.

6.1 La valutazione

Un primo ambito in cui pare evidente una possibilità di impiego pratico dell'indice di complessità morfologica riguarda la valutazione. L'ICM serve essenzialmente a valutare le produzioni linguistiche, a descriverle sotto un certo punto di vista in modo esplicito e rigoroso. Come in ogni forma di valutazione, occorre domandarsi preliminarmente quale è il suo scopo, cioè la rilevanza pratica per cui un insegnante dovrebbe essere interessato a misurare la complessità linguistica delle produzioni dei suoi alunni.

Pare evidente che i possibili impieghi dell'ICM non possano riguardare la valutazione sommativa. Che senso avrebbe attribuire un certo voto, o stabilire il superamento di un certo livello soglia, per chi produce testi con un indice di complessità morfologica superiore a un determinato valore? Per fare ciò, occorrerebbe che la complessità morfologica figurasse tra gli obiettivi di un syllabo didattico e certificatorio, con la specificazione dei diversi gradi intermedi corrispondenti a diversi punteggi, o livelli di una scala di valutazione. Questo pare problematico non tanto su un piano pratico, ma piuttosto teorico, di validità del costrutto. Si ricorderà che l'indice calcola la diversità formale degli esponenti flessivi, senza prendere in considerazione il loro valore semantico-funzionale né tantomeno la correttezza grammaticale o l'appropriatezza alla situazione comunicativa. Questi limiti, che sono stati intenzionalmente assunti per valide ragioni teoriche, rendono discutibile l'inserimento diretto, non mediato, dell'ICM in un syllabo linguistico-comunicativo, come quello che sottende i principali approcci glottodidattici contemporanei, compreso quello proposto dal QCER,⁴⁰ in cui la valutazione degli aspetti grammaticali non può prescindere da considerazioni sulla loro valenza comunicativa. Si potrebbero persino immaginare comportamenti opportunistici di candidati che usino la maggior gamma possibile di esponenti morfologici per ottenere un punteggio elevato su questo parametro, il tutto a prescindere dalla loro reale rilevanza comunicativa.

Se rimane dunque escluso un uso dell'ICM per la valutazione sommativa, questo diventa invece molto più rilevante come forma di valutazione formativa, nell'ottica della valutazione dell'interlingua come comprensione dei processi e analisi diagnostica. Lo studio di Brezina e Pallotti 2016 discusso in precedenza mostra che, in una lingua come l'italiano L2, la complessità morfologica cresce gradualmente con l'aumentare del livello di competenza linguistica. Pur ribadendo le note di cautela esposte pocanzi sul fatto che l'ICM non si può identificare come lo sviluppo della competenza morfologica tout court, esso rimane tuttavia un indicatore globale abbastanza affidabile dello sviluppo dei paradigmi verbali, specialmente per quanto riguarda l'introduzione del repertorio delle varie forme. In tal senso, può servire all'insegnante, insieme ad altre forme più qualitative e approfondite di analisi delle interlingue, per misurare i progressi degli allievi, identificando eventuali momenti di stallo o di 'slancio' nella sperimentazione delle numerose forme del verbo italiano. Esso consentirà di vedere, ad esempio, come nelle fasi iniziali il verbo si presenti in paradigmi piuttosto semplici, con la ripetizione di poche forme costanti sia per il limitato repertorio di tempi, aspetti e modi verbali, sia per la ben nota tendenza degli

⁴⁰ Cfr. Consiglio d'Europa 2001.

apprendenti a ridurre l'allomorfia: da quest'ultimo punto di vista, forme come *scendato* o *discutato* conterranno come ripetizioni del *type* morfologico *-ato*, presente anche in forme regolari come *parlato* e *cantato*, mentre le loro controparti standard *sceso* e *discusso*, presenti in varietà più avanzate, rappresentano diversi tipi morfologici, che fanno aumentare il valore di diversità dell'ICM. È in questo senso che l'ICM può contribuire a delineare il quadro evolutivo di un apprendente, nell'ottica della valutazione diagnostica.

6.2 L'insegnamento

L'ICM può trovare diverse applicazioni nell'ambito della didattica, in particolare all'interno di quelle attività definite come 'riflessione sulla lingua' o 'language awareness'. Tali attività, sebbene non abbiano lo scopo di insegnare direttamente nuove forme linguistiche o di consolidarne l'uso, hanno in ogni caso un ruolo importante nell'apprendimento di una seconda lingua, specialmente da parte di individui cognitivamente più maturi come gli adolescenti e gli adulti. Esse infatti promuovono un approccio di scoperta, di ricerca e verifica delle ipotesi, che responsabilizza gli apprendenti, dando loro un ruolo attivo e rendendoli di fatto dei ricercatori. Proviamo ora ad elencare alcune di queste attività, esemplificandole sull'italiano, ma tenendo presente che esse possono essere rilevanti anche per altre lingue.

7. Analizzare la complessità morfologica dei testi in italiano

Come si è visto nella presentazione del primo studio,⁴¹ anche all'interno di una stessa lingua la complessità morfologica varia a seconda dei testi; il secondo studio, d'altra parte, mostra come la padronanza di una maggiore complessità morfologica sia associata a livelli di competenza linguistica più alti. Quindi, l'ICM, insieme ad altri indici, come quelli che misurano la diversità lessicale o la lunghezza media delle strutture sintattiche, può dare un'indicazione di massima sulla varietà delle forme presenti in un testo, che può essere correlata alla varietà delle forme prodotte dagli apprendenti e alla loro capacità di elaborare e comprendere paradigmi verbali più o meno ampi. È chiaro che la corrispondenza non è diretta: si tratta della ben nota discrasia tra ampiezza del repertorio in comprensione e in produzione, ma anche del fatto che per comprendere un testo conta soprattutto il lessico, mentre la morfologia gioca un ruolo più secondario. Tuttavia, è altrettanto chiaro che la presenza di una vasta gamma di esponenti morfologici verbali può risultare difficile al lettore di livello medio-basso, e che in alcuni casi una loro mancata comprensione può portare a fraintendimenti simili a quelli prodotti dalla scarsa competenza lessicale.

Queste riflessioni possono essere svolte da un'insegnante che cerca di valutare la complessità dei testi da proporre agli alunni, ma anche dagli autori di libri di testo e materiali didattici, oltre che dai redattori di prove per i test linguistici. Tuttavia, si può immaginare anche un uso da parte degli alunni, che possono infatti calcolare autonomamente l'indice di complessità morfologica di diversi testi e riflettere sulle cause del suo aumento o della sua diminuzione. Essi noteranno ad esempio l'alta varietà di forme verbali presente nei testi narrativi, a causa dell'alternanza tra il presente dei dialoghi e il passato della narrazione, ma anche del fatto che questi passati sono spesso realizzati come passati remoti, con tutta l'allomorfia dei loro paradigmi, certamente più irregolari di altre aree della coniugazione verbale.

⁴¹ Cfr. Pallotti 2015b: 195-218.

8. Analizzare la complessità morfologica dei testi in interlingua

I testi scritti dagli alunni possono essere analizzati con lo strumento online per calcolarne la complessità morfologica. In tal modo, gli apprendenti possono diventare consapevoli dell'ampiezza dei loro repertori morfologici e delle forme che utilizzano man mano che progrediscono nell'apprendimento della L2. I risultati dell'analisi dell'ICM possono essere conservati nel portfolio delle loro produzioni, per documentare i progressi, un aspetto che stimola la motivazione. Inoltre, questa auto-osservazione della propria interlingua favorisce l'attenzione alle forme linguistiche, promuovendone l'acquisizione.⁴² Nelle fasi più basiche, come notato in precedenza, si risconterà la frequente ripetizione di pochi esponenti morfologici, e la conseguente regolarizzazione dei paradigmi contenenti allomorfi: gli apprendenti potranno riflettere sulle strategie che li hanno portati a produrre forme come *lui beva*, o *ha smettato*, non vedendole come semplici 'errori', deviazioni da una norma standard con una logica difficile da comprendere (che infatti spesso non c'è, o affonda le radici in varietà precedenti della lingua, o addirittura altre lingue), ma al contrario come forme che presentano una loro logica nella ricerca di regolarità, nonostante l'irregolarità di molte forme del verbo italiano. Insomma, gli apprendenti stessi dovrebbero rendersi conto dell'indipendenza di costrutti come complessità e accuratezza. Mettendo a confronto i valori dell'ICM nei loro testi e quelli nei testi scritti dai nativi, gli apprendenti potranno capire che l'acquisizione della morfologia in italiano consiste, tra le altre cose, nel padroneggiare un sistema complesso, ricco di forme diverse che spesso esprimono gli stessi significati. Ciò non implicherà automaticamente, come è ovvio, l'acquisizione delle nuove forme, ma favorirà la consapevolezza metalinguistica che, indirettamente, promuove il focus on form e dunque la ristrutturazione dell'interlingua.⁴³

9. Riflettere sui processi morfologici in italiano

L'operativizzazione dell'ICM poggia su una definizione di processo morfologico come trasformazione di una base di default in una forma flessa. Come si è detto, una simile definizione non ha la pretesa di essere valida sul piano psicologico, storico o della teoria linguistica, ma ha il pregio di essere esplicita, chiara e replicabile con le stesse modalità in diverse lingue.

Studenti più avanzati, sia nella loro conoscenza dell'italiano sia nella loro maturazione cognitiva e metalinguistica, potrebbero prendere questa definizione operativa di processo morfologico come punto di partenza per discutere formulazioni alternative. Ad esempio, potrebbero rendersi conto che in molti casi si tratta di due processi consecutivi, e non di uno unico: la definizione di esponente dell'ICM tratta una forma flessa come *parlavo* come se fosse l'aggiunta dell'esponente *-avo* alla base *parl-*. Tuttavia, è chiaro che quest'ultimo esponente può essere a sua volta scomposto in una vocale tematica *-a-*, in un suffisso tempo-aspettuale *-v-* che designa l'imperfetto e in un suffisso di persona *-o* che indica la prima persona singolare. O ancora: nei verbi irregolari la cui base cambia al participio passato e al passato remoto è ugualmente distinguibile un suffisso di persona che rimane costante al di là del mutare delle basi. Ad esempio, *pres-i*, *mis-i*, *diss-i* contengono tutti lo stesso suffisso di prima persona singolare *-i*, così come *pres-e*, *mis-e*, *diss-e* contengono quello di terza singolare *-e*. Dunque, si potrà ragionare sul fatto che, mentre i 'processi morfologici che trasformano le basi' *prend-*, *mett-* e *dic-* sono diversi, rimangono uguali i suffissi di persona, almeno in queste classi di verbi. Ma si potrà riflettere anche sulla validità di questa formulazione, un po' ingenua sul piano storico: è chiaro che molte di queste

⁴² Cfr. Ellis 2016: 405-428.

⁴³ Cfr. McLaughlin 1990: 113-128.

voci irregolari del verbo in italiano non ‘derivano’ da altre forme verbali dell’italiano, ma sono la continuazione di forme latine. Si può dunque aprire una discussione, e magari una ricerca, sull’etimologia delle forme del verbo italiano, che rende l’apprendimento di tutti questi allomorfi un processo di scoperta e analisi, piuttosto che una mera memorizzazione di liste di forme irregolari. O anche, senza sviluppare il piano diacronico, si potrà far notare che alcuni verbi italiani non hanno un solo tema, ma due, tre o più. Ad esempio, come si è detto, *prendere* ha due temi: *prend-*, che vale per quasi tutte le forme del paradigma, e *pres-*, che si usa per il participio passato *preso* e per la coniugazione di alcune persone del passato remoto. *Mettere* ne ha tre: *mett-* che vale per buona parte del paradigma, *mess-* che sta alla base del participio passato, e *mis-*, con cui si formano alcune persone del passato remoto, come *misi*, *mise*, *misero*. Anche in questo caso, una ricerca intelligente e scientificamente fondata su come rappresentare adeguatamente i processi morfologici rende il loro apprendimento più motivante, e l’analisi puramente formale del sito che calcola l’ICM può costituire, grazie alla sua semplicità, il punto di partenza per una simile riflessione.

10. Correggere le analisi automatiche

Lo strumento online per l’analisi della complessità morfologica, come tutti gli strumenti automatici, a volte commette degli errori (attualmente rappresentano meno del 5% delle forme). Questi dipendono il più delle volte dalla lemmatizzazione e analisi in parti del discorso di Treetagger, che può assegnare i verbi a una diversa classe di parole (vedendoli, ad esempio, come nomi, o aggettivi), oppure male interpretare alcune forme verbali. Correggere questi errori può diventare un’attività didattica utile e divertente. Gli apprendenti si trovano, per una volta, dalla parte di chi cerca l’errore e non di chi se lo fa cercare dall’insegnante: di nuovo, ciò li pone in un atteggiamento attivo, critico, di scoperta, che favorisce l’attenzione alla forma e la consapevolezza metalinguistica. In altri termini, riflettere sui possibili errori dell’analisi automatica implica consultare grammatiche di riferimento, discutere a coppie o in gruppi le ipotesi circa la morfologia, su cosa è corretto e cosa no, chiedere eventualmente aiuto all’insegnante su queste ipotesi: tutte attività di *focus on form* che lasciano tracce durature nella memoria.⁴⁴ Inoltre, questa riflessione porta gli alunni a utilizzare con maggiore sicurezza categorie di descrizione della lingua, come i termini per le diverse classi di parole (nomi, verbi, aggettivi, avverbi ...) e le loro forme (imperativo, gerundio, participio ...).

11. Conclusioni

In questo articolo abbiamo presentato un nuovo modo per misurare la complessità morfologica dei testi, l’Indice di Complessità Morfologica (ICM), applicandolo a due contesti di ricerca. Da un lato, in una prospettiva tipologico-comparativa, si è mostrato come diverse lingue abbiano diversi livelli di complessità morfologica; tuttavia quest’ultima varia anche, in modo sostanziale, tra diversi tipi di testo. In una prospettiva acquisizionale, invece, si è visto come, in una lingua morfologicamente ricca come l’italiano, la complessità morfologica aumenti in parallelo con lo sviluppo delle competenze nella L2.

Nonostante l’ICM sia stato sviluppato in primo luogo per obiettivi di ricerca, esso può essere utilizzato anche per finalità didattiche, come si è cercato di mostrare nella seconda parte dell’articolo. Il suo uso più naturale, in tal senso, è come contributo a una valutazione diagnostica, formativa, che prenda atto dell’estendersi dei paradigmi morfologici degli studenti, registrandone l’ampiezza e il tipo di forme che man mano

⁴⁴ Cfr. Ellis 2016: 405-428.

si sviluppano, non sempre in modo standard, ma seguendo spesso logiche interlinguistiche comunque razionali e funzionali.

D'altra parte, sono possibili anche attività didattiche che, prendendo le mosse dalle analisi, automatiche e non, della complessità morfologica tramite l'ICM, portano a riflettere criticamente sulla morfologia dell'italiano nativo e non nativo. Si tratta di attività adatte soprattutto a un pubblico di adolescenti o adulti, di livello intermedio e avanzato, ma che promuovono una coscienza metalinguistica che rende gli apprendenti più autonomi, motivati e consapevoli. In questo ambito diverse ricerche hanno dimostrato come questi atteggiamenti siano favorevoli allo sviluppo della L2. Queste attività di riflessione sulla forma non sono da vedere come estranee, e tantomeno antitetiche, rispetto a un approccio comunicativo. Al contrario, le attività in cui gli apprendenti discutono sulle forme linguistiche, sulle loro ipotesi e convinzioni in merito, sono tra le più genuinamente comunicative che si possano immaginare: in un'aula di lingue, cosa c'è di più autentico di una discussione sulla lingua?

Bibliografia

- Bernardini, P. & J. Granfeldt. 2014. 'Linguistic complexity and cross-linguistic variation in L2 English, L3 French and L4 Italian'. Relazione al ReN workshop, AILA Conference, Brisbane, August, 10-15.
- Brezina, V. & G. Pallotti. 2016. 'Morphological complexity in written L2 texts'. *Second Language Research*, doi:10.1177/0267658316643125, July 2016.
- Brumfit, C. J. & K. Johnson. 1979. *The communicative approach to language teaching*. Oxford: Oxford University Press.
- Bulté, B. & A. Housen. 2012. 'Defining and operationalising L2 complexity'. In: *Dimensions of L2 performance and proficiency - Investigating complexity, accuracy and fluency in SLA*, a cura di A. Housen, F. Kuiken & I. Vedder, 21-46. Amsterdam/Philadelphia: Benjamins.
- Chomsky N. 1992. *What Uncle Sam really wants*. Berkeley, CA: Odonian Press.
- Consiglio d'Europa. 2001. *Quadro comune europeo di riferimento per le lingue: Apprendimento, insegnamento, valutazione*. Milano: La Nuova Italia.
- De Clercq, B. & A. Housen. 2017. 'A cross-linguistic perspective on syntactic complexity in L2 development: Syntactic elaboration and diversity'. *The Modern Language Journal*, 101(2): 315-334.
- De Clercq, B. 2015. 'The development of lexical complexity in second language acquisition. A cross-linguistic study of L2 French and English'. *Eurosla Yearbook*, 15: 69-94.
- Devoto, G. & G.C. Oli. 2000. *Dizionario della lingua italiana*. Firenze: Le Monnier.
- Eco, U. 1992. [1996]. *La ricerca della lingua perfetta nella cultura europea*. Roma: Laterza.
- Ellis, R. 2016. 'Focus on form: A critical review'. *Language Teaching Research*, 20(3): 405-428.
- Gyllstad, H., J. Granfeldt, P. Bernardini & M. Källkvist. 2014. 'Linguistic correlates to communicative proficiency levels of the CEFR: The case of syntactic complexity in written L2 English, L3 French and L4 Italian'. *Eurosla Yearbook*, 14: 1-30.
- Haspelmath, M. & A. Sims. 2010. *Understanding morphology, 2nd ed.* London: Hodder Education.
- Horst, M. & L. Collins. 2006. 'From faible to strong: How does their vocabulary grow?' *Canadian Modern Language Review*, 63(1): 83-106.
- Housen, A. & F. Kuiken. 2009. 'Complexity, accuracy and fluency in second language acquisition'. *Applied Linguistics*, 30(4): 461-473.

- Housen, A., F. Kuiken & I. Vedder. cur. 2012. *Dimensions of L2 performance and proficiency. Complexity, accuracy, and fluency in SLA*. Amsterdam/Philadelphia: John Benjamins.
- Hunt, K. W. 1965. 'Grammatical structures written at three grade levels'. *NCTE Research Report* No. 3.
- Johnson, W. 1944. 'Studies in language behavior: I. A program of research'. *Psychological Monographs*, 56(2): 1-15.
- Klein-Braley, C. & U. Ratz. 1982. 'Der C-Test: ein neuer Ansatz zur Messung allgemeiner Sprachbeherrschung'. *AKS-Rundbrief*, 4: 23-37.
- Kuiken, F. & I. Vedder. 2014. 'Rating written performance: What do raters do and why?'. *Language Testing*, 31(3): 329-348.
- Kuiken F., I. Vedder & R. Gilabert. 2010. 'Communicative adequacy and linguistic complexity in L2 writing'. In: *Communicative proficiency and linguistic development*, a cura di I. Bartning, M. Martin, & I. Vedder, 81-99. Amsterdam: Eurosla Monograph series.
- Kortmann, B. & B. Szmrecsanyi. cur. 2012. *Linguistic complexity: Second language acquisition, indigenization, contact*. Berlin: de Gruyter.
- Long, M. 2015. *Second language acquisition and task-based language teaching*. Malden, MA: Wiley-Blackwell.
- Malvern D., B. Richards, N. Chipere & P. Durán. 2004. *Lexical diversity and language development: Quantification and assessment*. Basingstoke: Palgrave MacMillan.
- McLaughlin, B. 1990. Restructuring. *Applied Linguistics*, 11: 113-128.
- McWhorter, J. 2001. 'The world's simplest grammars are creole grammars'. *Linguistic Typology*, 5: 125-166.
- McWhorter, J. 2007. *Language Interrupted*. Oxford: Oxford University Press.
- Michel, M. 2017. 'Complexity, Accuracy and Fluency in L2 production'. In: *Routledge Handbook of Instructed Second Language Acquisition*, a cura di S. Loewen & M. Sato, 50-68. London: Routledge.
- Miestamo, M., K. Sinnemäki & F. Karlsson. cur. 2008. *Language complexity: typology, contact, change*. Amsterdam: Benjamins.
- Norris, J. & L. Ortega. 2009. 'Measurement for understanding: An organic approach to investigating complexity, accuracy, and fluency in SLA'. *Applied Linguistics*, 30: 555-578.
- Nuzzo, E. & P. Gauci. 2012. *Insegnare la pragmatica in italiano L2*. Roma: Carocci.
- Pallotti, G. 1999. 'Relatività linguistica e traduzione'. *Versus. Quaderni di studi semiotici*, 82: 109-138.
- Pallotti, G. 2009. 'CAF: Defining, refining and differentiating constructs'. *Applied Linguistics*, 30: 590-601.
- Pallotti, G. 2015a. 'A simple view of linguistic complexity'. *Second Language Research*, 31: 117-134.
- Pallotti, G. 2015b. 'Una nuova misura della complessità linguistica: L'Indice di Complessità Morfologica (ICM)'. *Rivista Italiana di Linguistica Applicata*, 2: 195-218.
- Pallotti, G. & S. Ferrari. 2008. 'La variabilità situazionale dell'interlingua: Implicazioni per la ricerca acquisizionale e il testing linguistico'. In: *Competenze lessicali e discorsive nell'acquisizione di lingue seconde*, a cura di G. Bernini, L. Spreafico & A. Valentini. Perugia: Guerra.
- Pallotti, G. 2017. 'Applying the interlanguage approach to language teaching'. *International review of applied linguistics in language teaching*, 55: 393-412.
- Sampson, G. 2009. 'A linguistic axiom challenged'. In: *Language Complexity as an Evolving Variable*, a cura di G. Sampson, D. Gil & P. Trudgill, 1-18. Oxford: Oxford University Press.

- Sampson, G., D. Gil & P. Trudgill, cur. 2009. *Language complexity as an evolving variable*. Oxford: Oxford University Press.
- Schmid, H. 1994. 'Probabilistic part-of-speech tagging using decision trees'. *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK.
- Skehan, P. 1998. *A cognitive approach to language learning*. Oxford: Oxford University Press.
- Trudgill, P. 1989. 'Contact and isolation in linguistic change'. In: *Language Change: contributions to the study of its causes*, a cura di L. Breivik & E. Jahr, 227-237. Mouton de Gruyter, Berlin.
- Wolfe-Quintero, K., S. Inagaki, & H. Y. Kim. 1998. *Second language development in writing: Measures of fluency, accuracy, & complexity*. Honolulu: University of Hawaii Press.

Parole chiave

Complessità linguistica, morfologia, coniugazione verbale, attenzione alla forma, language awareness

Gabriele Pallotti insegna Didattica delle Lingue moderne all'Università di Modena e Reggio Emilia. Le sue ricerche riguardano i rapporti tra socializzazione e apprendimento della seconda lingua, analisi dell'interlingua, metodologia della ricerca in linguistica applicata. Ha coordinato progetti nazionali e locali di formazione degli insegnanti e di sperimentazione educativa e ha tenuto numerose conferenze e lezioni in Italia e all'estero. Fa parte del direttivo della rete europea SLATE (Second Language Acquisition and Testing in Europe) e della European Second Language Association (Eurosla), per cui è responsabile della serie di monografie Eurosla Studies. È autore e curatore di diversi volumi, tra cui *La seconda lingua* (Bompiani), *Scrivere per comunicare* (Bompiani), *Imparare e insegnare l'italiano come seconda lingua* (Bonacci), *L2 Learning as social practice* (University of Hawai'i Press), *Che storia!* (Loescher).

Dipartimento di Educazione e Scienze umane
Università di Modena e Reggio Emilia
Viale Allegri 9
42121 Reggio Emilia (Italia)
gabriele.pallotti@unimore.it

SUMMARY

Morphological complexity

Research and teaching perspectives

The article discusses a new construct and its operationalization, morphological complexity (MC). After reviewing the current literature on complexity in linguistic typology and second language acquisition research, MC is theoretically defined in terms of the range and variety of morphological processes (also called exponents) instantiated in a given text. This theoretical definition is then operationalized with a procedure (freely available online) which first identifies exponents as morphological operations on a lexical base, and then calculates the average diversity within and across samples of 10 exponents. At present, MC is calculated on verb inflections only. A first study shows that MC varies in different texts and different languages (Italian, German and English). In a second study, MC is calculated in learners of L2 Italian, and is shown to be positively correlated to general linguistic competence, as measured with a C-test, and lexical and syntactic complexity. Finally, a number of possible practical uses of the construct in a language classroom are suggested. These include the formative evaluation of learners' morphological repertoires, and awareness-raising activities that stimulate learners to reflect on Italian's verbal morphology, its structure, history, and theoretical description.