

ATTIVITÀ METALINGUISTICHE SU SUPPORTO INFORMATICO PER LO SVILUPPO DELLA COMPETENZA ORTOGRAFICA DI APPRENDENTI L'ITALIANO L2

Jacopo Saturno¹

1. QUADRO TEORICO

Il presente studio è dedicato al problema della competenza ortografica nell'italiano L2 di un gruppo di studenti polacchi principianti, per potenziare la quale si propone un'attività metalinguistica basata sulle espressioni regolari, le quali costituiscono uno strumento informatico molto comune nell'ambito dell'umanistica digitale. Ci si concentrerà in particolare sulla complessa relazione tra la lettera <c> e i numerosi fonemi che concorre a rappresentare in quanto componente di grafemi complessi, quali <ch>, <sci> etc. Gli errori di ortografia in italiano sono stati oggetto di indagine prevalentemente nell'ambito della diagnosi della dislessia, cui spesso si accompagna anche la disortografia (Bozzo *et al.*, 2000; Angelelli *et al.*, 2008; Tressoldi *et al.*, 2013). Alla luce degli obiettivi del presente lavoro, il quale prende in considerazione il caso di un gruppo di apprendenti di italiano L2 presumibilmente non dislessici, pare opportuno adottare la classificazione di Tressoldi *et al.* (2013), nell'ambito della quale tutti i potenziali errori qui considerati appartengono alla categoria degli errori “fonologici”, in cui risulta distorta l'associazione tra fonemi e grafemi.

Fondamentale per l'analisi delle abilità di lettura degli apprendenti è lo studio della relazione tra grafemi e fonemi. Limitandosi al solo caso della scrittura alfabetica (si veda Ellis *et al.*, 2004 per i sistemi di altro tipo), un utile punto di riferimento è costituito dal principio fonologico, per il quale a ogni grafema corrisponde un solo fonema e viceversa, anche a scapito della coerenza nella rappresentazione di morfemi condivisi tra parole di diversa pronuncia. Per esempio, l'ortografia del bielorusso registra la differente pronuncia di diverse forme di parola del medesimo lessema, rispecchiando così la regola fonologica per cui il fonema /o/ in posizione atona tende a essere pronunciato [Λ], es. *vada* [vΛ'da] “acqua: NOM.SG” vs. *vody* [ˈvɔdʲ] “acqua: NOM/ACC.PL”. Normalmente però nelle lingue storico-naturali si rileva un certo scostamento da tale principio, con esiti più o meno prevedibili. In russo, per esempio, le consonanti sonore che in posizione finale di parola si desonorizzano sono rappresentate graficamente dal medesimo grafema che le esprime in ogni altro contesto, il quale così finisce per corrispondere a due pronuncie diverse, es. *ryba* [ˈrʲɪbɐ] ‘pesce: NOM.SG’ vs. *ryb* [rʲɪp] ‘pesce: GEN.PL’. Tale comportamento è però sistematico e prevedibile, così che la possibilità di leggere e pronunciare parole non note non risulta pregiudicata. Al contrario, nella medesima lingua il grafema <o> può corrispondere ai suoni [o, Λ, ə] a seconda della sua posizione rispetto alla sillaba accentata, mentre i suoni [Λ, ə] possono corrispondere ai grafemi <a> oppure <o> a seconda dell'etimologia della parola. Data una parola russa, dunque, non è univocamente possibile

¹ Università degli Studi di Bergamo.

ricostruirne la pronuncia a partire dalla sua rappresentazione ortografica, né trascriverla ortograficamente a partire dalla pronuncia.

La relazione tra fonemi e grafemi è alla base dell'ipotesi della trasparenza ortografica (*orthographic depth*), la quale sostiene che esistano due principali meccanismi di decodifica della parola grafica. Il primo prevede il progressivo assemblaggio della forma fonica mediante l'associazione di ciascun grafema al fonema corrispondente. Nel secondo invece il lessico mentale del lettore è esplorato alla ricerca di un elemento che corrisponda alla forma grafica complessiva della parola o almeno dei morfemi che la compongono (si veda Albano Leoni, 2009, per un parallelo nell'ambito della fonologia). Il primo meccanismo sarebbe dominante nelle ortografie trasparenti, mentre il secondo si rivelerebbe più efficace in quelle opache, per quanto nella sua versione moderata l'ipotesi postuli che entrambi i meccanismi contribuiscano al processo di lettura, ma in proporzioni variabili a seconda del sistema ortografico della lingua (Katz, Frost, 1992). L'apprendimento della lettura risulta più rapido nel caso delle ortografie trasparenti (Cossu *et al.*, 1988; Seymour *et al.*, 2003), tanto che nel caso dell'ebraico, la cui scrittura di tipo abjad può essere integrata da segni diacritici rappresentanti le vocali, per l'insegnamento della lettura è preferita proprio questa versione più trasparente. Ellis *et al.* (2004) riassumono alcuni risultati a sostegno dell'ipotesi della trasparenza ortografica. In primo luogo, la capacità di leggere correttamente le non-parole è notevolmente superiore nel caso delle ortografie trasparenti. Similmente, in queste lingue gli errori consistono spesso nella produzione di non-parole, laddove nelle ortografie opache vengono spesso prodotte parole diverse da quelle richieste, ma esistenti nel lessico della lingua. Infine, la lunghezza di una parola in termini di caratteri influenza i tempi di lettura in misura nettamente maggiore nel caso delle ortografie trasparenti. Tutto ciò porta a concludere che le ortografie trasparenti inducono il lettore ad assemblare la forma fonica delle parole grafema per grafema, laddove in quelle opache la lettura è il frutto del riconoscimento olistico dell'immagine grafica corrispondente alla parola. Quest'ultima strategia evidentemente prevede che le parole siano già note al lettore nella loro rappresentazione tanto fonologica, quanto ortografica.

Diversi studi mostrano che il tipo di ortografia della L1 può influenzare l'approccio adottato nella lettura di eventuali lingue straniere apprese in età adulta (Bhide, 2015; van Daal and Wass, 2017). Mentre è bene tenere presente questa variabile, essa non sembra direttamente rilevante per il presente studio, in cui le due lingue in contatto (italiano e polacco) condividono un'ortografia pressoché fonemica, per quanto ricca di grafemi composti, per effetto dei quali la relazione tra lettere e fonemi è spesso di uno-a molti.

Gli studi sugli errori ortografici in italiano L2 non sono particolarmente numerosi, specialmente per quanto riguarda gli apprendenti di origine slava. La sezione del *corpus* VALICO dell'italiano L2 scritto (Corino, Marelli, 2009) dedicata agli apprendenti polacchi non contiene errori di ortografia particolarmente rilevanti per questo lavoro, se non occasionali occorrenze nella grafia del grafema /k/ quali *que* (che), *chosa* (cosa), *mechanico* (meccanico), *scerci* (scherzi) etc. In uno studio sull'italiano scritto di 29 maturandi sloveni (ai quali è richiesta una competenza di livello B1), Bažec (2019) riporta che gli errori rilevanti ai fini del presente studio ammontano al 13,5% dei 1010 errori ortografici trovati nel *corpus*. I più frequenti riguardano il fonema /k/, per rappresentare il quale è talvolta erroneamente utilizzata la lettera <h> avanti consonante o vocale posteriore, es. *barcha*, *chredo*; dall'altra parte la medesima lettera è spesso omessa avanti vocali anteriori, es. **qualce*, **cilometri*. Simili errori prevedono l'aggiunta o omissione del diacritico <i> per rappresentare i fonemi /ʃ/, es. *lascate* (lasciate), e /ʒ/, es. *piacono* (piacciono). In alcuni casi è ipotizzabile un influsso dell'ortografia della L1, come nella rappresentazione dei fonemi /ʦ/, /z/ o /k/ (in sloveno come in polacco rispettivamente <c, z, k>), es. *abbastanca*, *misteriozi*, *kome*. Allontanandosi di poco dall'area slava, simili errori si riscontrano

nell'italiano L2 di studenti lituani, ma solo ai livelli A1 o al più A2, il che spiegherebbe la loro scarsità nei testi relativamente complessi del *corpus* VALICO (Gataveckaitė, 2013: 142).

Vale infine la pena di spendere qualche parola sulla dinamica del processo di apprendimento della lettura. I principali modelli (Frith, 1986; Ehri, 2005) prevedono un'acquisizione per fasi, in cui l'apprendente (normalmente un bambino) dapprima considera le parole come qualunque altra unità grafica (icone, immagini), poi impara le corrispondenze tra lettere e suoni, poi ancora quelle tra grafemi e suoni, infine crea un lemma specifico per ogni unità di livello superiore (morfemi, intere parole) in un lessico ortografico mentale. Chiaramente tali fasi di sviluppo procedono di pari passo con l'esperienza nella lingua bersaglio: l'ultima fase in particolare non può prescindere da una conoscenza anche semantica degli elementi della lingua. Nell'acquisizione di L2, tale processo graduale e parzialmente implicito è sostituito con l'apprendimento esplicito e rapido di una serie di regole metalinguistiche riguardanti la relazione tra fonemi e grafemi. Il presente lavoro insiste proprio su questo momento trascurato eppure cruciale dell'apprendimento di L2 in ambito accademico, proponendosi di verificare l'efficacia di un'attività metalinguistica volta a esplicitare e sistematizzare la relazione tra le unità fonologiche della lingua e la loro rappresentazione ortografica.

2. UN ESPERIMENTO DIDATTICO

Si descrive qui una metodologia didattica sperimentale volta a facilitare l'identificazione di quei tratti del sistema ortografico dell'italiano che maggiormente lo allontanano dal principio fonologico. L'esercizio consiste nell'elaborazione di una serie di operazioni di "trova e sostituisci", per mezzo delle quali un testo ortografico è progressivamente trasformato nella sua trascrizione fonologica². La Tabella 1 riassume i passaggi necessari per trascrivere la parola polacca *Szczecin* /'ʃʃɛtɛin/ "Stettino". Il ciclo 1 trova tutte le occorrenze di <sz> e le sostituisce con [ʃ], il secondo fa lo stesso con <cz> e [tʃ], rispettivamente, e così via. La pronuncia dei grafemi <sz> e <cz> non dipende dal contesto ortografico, così che è sufficiente cercare il gruppo formato dalle due lettere. Lo stesso non vale però per il grafema <ci>, la cui pronuncia differisce a seconda del fatto che il nesso sia seguito oppure no da una vocale. Si rende dunque necessario utilizzare l'espressione regolare "(?![aeou])", la quale specifica che il nesso <ci> non deve essere seguito dalle lettere <a, e, o, u>.

Tabella 1. *Trascrizione della parola polacca Szczecin "Stettino" mediante "trova e sostituisci"*

ciclo	trova	sostituisci	stringa
0	-	-	Szczecin
1	sz	ʃ	ʃczecin
2	cz	tʃ	ʃtʃecin
3	ci(?![aeou])	tei	ʃtʃetein

Le espressioni regolari (*regex*) sono funzioni che permettono di specificare una stringa di ricerca in cui alcuni elementi hanno un valore astratto, piuttosto che corrispondere a un carattere specifico. Per esempio, l'operatore <.> corrisponde a qualsiasi carattere, così

² Si segnala che in tutto il lavoro non saranno prese in considerazione la posizione dell'accento tonico né la lunghezza vocalica e consonantica.

che cercando l'espressione regolare “c.nto” in un *corpus* di parole italiane si troveranno occorrenze come *canto, cento, cinto, conto*.

Nel presente esperimento si farà uso di una selezione molto limitata di espressioni regolari, descritte nella Tabella 2 ed esemplificate sulla base degli esempi polacchi in (1).

(1) *goń, kod, kot, kota, kotowi, koń, koty, kotom, los*³

Tabella 2. *Espressioni regolari utilizzate nell'esperimento*

regex	significato	stringa	risultato
[xyz]	uno qualsiasi dei caratteri inclusi tra parentesi	kot[ay]	<i>kota</i> e <i>koty</i> ma non <i>koto</i> in <i>kotowi</i> etc.
(?<!y)x	x non preceduto da y, senza selezionare il carattere che precede x	(?<!k)	<i>o</i> in <i>kotowi, goń</i> e <i>los</i> , ma non <i>kot</i> etc.
x(?!y)	x non seguito da y, senza selezionare il carattere che segue x	ko(?!t)	<i>ko</i> in <i>kod</i> e <i>koń</i> , ma non <i>kot</i> etc.

Combinando gli operatori è possibile cercare specifici grafemi, preceduti o seguiti (o viceversa, non preceduti o non seguiti) da altri grafemi. L'espressione “(?<!s)ci(?=[aeou])”, per esempio, permette di trovare in un testo tutte le occorrenze di <ci> non precedute da <s> e seguite da una lettera a scelta tra <a, e, o, u>, senza che queste ultime risultino però parte della selezione.

L'identificazione delle espressioni regolari necessarie per un sistema di trascrizione fonologico non è un compito banale. Oltre alla conoscenza di questo strumento, infatti, è necessaria una visione complessiva dell'ortografia della lingua bersaglio. È inoltre molto utile una frequente verifica empirica della correttezza del sistema elaborato. Si supponga di voler trascrivere fonologicamente la parola *elogio*. Molte delle lettere che la compongono non corrispondono a un unico carattere di un alfabeto fonetico, caratterizzato cioè da una relazione biunivoca tra simboli e suoni (es. IPA), in quanto la loro pronuncia varia a seconda del contesto grafico, cioè del fatto che concorrano oppure no a comporre un grafema composto di più lettere. La lettera <g>, ad esempio, può rappresentare i suoni /g/, /ŋ/, /ʎ/ oppure /dʒ/: quest'ultimo in particolare è quello corrispondente al digrafo <gi>, si direbbe intuitivamente. Per trascrivere adeguatamente questo suono sembrerebbe allora sufficiente cercare nel testo la stringa “gi” e sostituirla con /dʒ/, ottenendo /e'loʒo/. L'inadeguatezza di questa soluzione appare però subito evidente se la si applica alla parola *elogi*, che risulterebbe trascritta /e'loʒ/. In effetti la pronuncia del nesso <gi> dipende dal fatto che esso sia seguito da una vocale (come in *elogio*) oppure si trovi in fine di parola (es. *elogi*), o ancora sia seguito da una consonante (es. *giro*). Per trovare e sostituire correttamente le diverse occorrenze di <gi>, dunque, occorre specificare due diverse stringhe di ricerca. La prima, “gi(?=[aeou])”, identificherà i casi in cui il nesso è seguito da una vocale, es. *elogio*, che saranno sostituiti da /dʒ/, es. /e'loʒo/; l'altra, “gi(?![aeou])”, troverà invece i casi in cui non lo è, es. *elogi*, che saranno sostituiti da /dʒi/, es. /e'loʒi/. Si noti in ogni caso che questa è solo una delle possibili soluzioni al problema della trascrizione di questo grafema.

Nel presente studio si è chiesto ai partecipanti di elaborare una lista di espressioni regolari che descrivano le regole di lettura della lettera <c> in italiano, nella convinzione che l'approfondito lavoro metalinguistico necessario a questo scopo avrebbe influito positivamente sulla loro competenza ortografica.

³ Traduzioni: “insecure:IMP.2SG”, “codice:NOM/ACC.SG”, “gatto:NOM.SG”, “gatto:GEN/ACC.SG”, “gatto:DAT.SG”, “cavallo:NOM.SG”, “gatti:NOM/ACC”, “gatti:DAT”, “destino:NOM/ACC.SG”.

3. STRUTTURA BERSAGLIO

In senso stretto, l'unica vera fonte di imprevedibilità nell'ortografia italiana è rappresentata dal fatto che l'accento non è rappresentato graficamente se non nelle parole ossitone, es. *però*, oppure per distinguere tra loro particolari coppie di omofoni, es. *ancora* /'ankora/ vs. *ancóra* /an'kora/. A ciò si lega anche l'interpretazione di alcuni gruppi di lettere in cui compare la lettera <i>, la quale a seconda che sia portatrice di accento oppure no dovrà rispettivamente essere interpretata come vocale oppure come diacritico, es. *scia* /'ʃi:a/ vs. *sciabola* /'ʃabola/. In tutti gli altri casi, a ciascun fonema corrisponde un solo grafema, inteso come un simbolo formato di una o più lettere inserite in un dato contesto grafico, e viceversa. Lo stesso però non vale per le singole lettere, le quali spesso concorrono a formare grafemi composti, cioè formati di più lettere, le quali a loro volta possono ricorrere in diversi grafemi, come la lettera <g> che ricorre nei grafemi <gh>, <gl>, <gn> etc. Le singole lettere, dunque, non sempre corrispondono univocamente a un fonema, così che, per quanto l'ortografia dell'italiano permetta di risalire quasi senza ambiguità alla pronuncia di una parola, la sua trasparenza è pur sempre minore rispetto a quella un'ortografia puramente fonemica, rappresentata ad esempio da un alfabeto fonetico.

Particolarmente rappresentativo a questo riguardo è il caso della lettera <c> (Tabella 3). Indipendentemente dall'eventuale grafema precedente, essa corrisponde al suono /k/ (regola 1) quando è seguita da un'altra consonante (es. *cratere*), da <h> (es. *chilo*) oppure da <a, o, u> (es. *cane*). In italiano, cioè è equivalente a dire che <c> non deve essere seguita da <e, i>. È pronunciata invece /tʃ/ (regola 2) quando non è preceduta da <s> ed è seguita da <e>, es. *cena* (regola 2c), oppure <i>. In quest'ultimo caso occorre specificare anche se il gruppo <ci> è seguito da un'altra vocale oppure no. Nel primo caso infatti (regola 2a) la lettera <i> ha unicamente valore di diacritico, in quanto non è rappresentato nella pronuncia, es. *ciotola*, altrimenti (regola 2b) esprime anche il suono /i/, es. *cinema*, *scorci*. La medesima logica si applica anche alla regola 3, in cui il gruppo <ci> preceduto da <s> corrisponde tuttavia al suono /ʃ/.

Tabella 3. *Regole di pronuncia della lettera <c> in italiano*

regola	precedente	seguinte	IPA	regex	esempi
1	irrelevante	≠ i,e ⁴	/k/	c(?![ie])	<i>cane, credo, chilo</i>
2a	≠ s	i seguita da a, e, o, u	/tʃ/	(?<!s)ci(?=[aeou])	<i>ciao, ciotola</i>
2b		i non seguita da a, e, o, u	/tʃi/	(?<!s)ci(?![aeou])	<i>cinema, scorci</i>
2c		e	/tʃe/	(?<!s)ce	<i>cena</i>
3a	s	i seguita da a, e, o, u	/ʃ/	sci(?=[aeou])	<i>sciame, sciocco</i>
3b		i non seguita da a, e, o, u	/ʃi/	sci(?![aeou])	<i>pesci, scimmia</i>
3c		e	/ʃe/	sce	<i>scena</i>

Si noti che non necessariamente l'interpretazione dell'ortografia italiana in termini di espressioni regolari ne ricalca la tradizionale descrizione glottodidattica. La regola 1, per esempio, privilegia la sinteticità rispetto all'esplicitezza, riunendo in una singola

⁴ In questo studio non si prendono in considerazione le varianti accentate delle vocali (<è, é, ì, ò, à, ù>).

espressione i casi in cui <c> è seguita da consonante (es. *credo*), vocale non anteriore (es. *cane, cono, barracuda*) e <h> (es. *chela, chilo*).

L'ortografia polacca presenta alcuni punti di contatto con quella italiana (Tabella 4). Si segnala in particolare una regola pressoché identica all'equivalente italiano, per cui il gruppo <ci> è pronunciato /tʃ/ se seguito da una vocale, /tʃi/ se viceversa il gruppo si trova in fine di parola o è seguito da una consonante (rispettivamente regole 3a e 3b). I fonemi corrispondenti (/tʃ/ vs. /tʃi/) differiscono solo in termini di luogo di articolazione, tanto da essere spesso confusi nelle varietà di apprendimento del polacco L2 (Shoemaker, 2015).

Tabella 4. Regole di pronuncia della lettera <c> in polacco

regola	seguinte	IPA	regex	esempi
1	irrilevante	/tʃ/	cz	<i>kurczak</i> /'kurʃak/ 'pollo'
2		/x/	ch	<i>mucha</i> /'mucha/ 'mosca'
3a	i non seguita da a, e, o, u	/tʃ/	ci(?=[aeou])	<i>bocian</i> /'botʃan/ 'cicogna'
3b	i seguita da a, e, o, u	/tʃi/	ci(?![aeou])	<i>dzieci</i> /'dʒetʃi/ 'bambini', <i>cichy</i> /'tʃixi/ 'silenzioso'
4	≠ z, h, i	/tʃ/	c(?![zhi])	<i>co</i> /tʃo/ 'cosa?', <i>cena</i> /'tʃena/ 'prezzo'

Vari grafemi composti dalla lettera <c> compaiono anche in altre lingue straniere con cui uno studente polacco di italianistica potrebbe verosimilmente entrare in contatto, come lo spagnolo (es. *chico* /'tʃiko/ 'piccolo') e l'inglese (es. *cherry* /'tʃeɪ/ 'ciliegia'), in cui il grafema esprime il suono /tʃ/, nonché il tedesco (es. *dach* /dax/ 'tetto').

4. DOMANDE DI RICERCA E IPOTESI

Lo studio si propone di verificare se e in quale misura la competenza ortografica di un gruppo di apprendenti principianti possa trarre giovamento da un'attività fortemente metalinguistica e svolta mediante mezzi informatici. In realtà, le regole relative alla struttura bersaglio avrebbero dovuto essere note ai partecipanti sin dalla prima lezione di lingua italiana. Lo studio intende però verificare l'ipotesi per cui esprimere una regola in termini di espressioni regolari richiederebbe una sua più profonda assimilazione rispetto a una descrizione verbale: diversa cosa sarebbe da un lato dire "il gruppo <scia> si legge /ʃa/", dall'altro associare le pronunce /ʃ/ e /ʃi/ alle due espressioni regolari "sci(?=[aeou])" e "sci(?![aeou])". Le domande di ricerca si possono dunque sintetizzare nelle formulazioni seguenti:

DR1: l'esercizio metalinguistico proposto influisce positivamente sulle abilità di decodifica ortografica dei partecipanti?

IP1: Si ipotizza che il lavoro metalinguistico necessario a definire le espressioni regolari possa esercitare un effetto positivo sulle capacità di lettura degli apprendenti, le quali nello stadio iniziale di acquisizione qui considerato sono verosimilmente fondate su un processo di assemblaggio fonologico.

DR2: l'eventuale miglioramento nei punteggi attribuibile all'esercizio varia a seconda della specifica struttura bersaglio considerata?

IP2: Da un lato, l'effetto positivo dell'esercizio dovrebbe risultare massimo nel caso di regole ortografiche molto specifiche, nelle quali cioè compaia in modo esplicito un numero limitato di lettere. L'espressione "sci(?=[aeou])", per esempio, esplicita tutti i potenziali elementi del contesto ortografico, laddove "sci(?![aeou])" esplicita le lettere che *non* ne possono fare parte, lasciando quindi aperto il novero (potenzialmente numeroso) di elementi che possono occupare quella posizione. Dall'altro lato, anche in assenza di un effetto positivo dell'esercizio, alcuni grafemi potrebbero risultare più accessibili per effetto dell'interferenza interlinguistica. Le varie pronunce del grafema <ci> a seconda del contesto ortografico sono molto simili tra italiano e polacco, e perciò potenzialmente assimilabili (si vedano Shoemaker, 2015; Best, 1995; Flege, 1995 per un parallelo nell'ambito della fonologia). È dunque verosimile che questi grafemi mostrino punteggi di accuratezza superiori. Dall'altra parte, il grafema <ch> è presente con un diverso valore fonetico in diverse lingue europee, compresa la L1 dei partecipanti, così che non è da escludere la presenza di fenomeni di interferenza negativa. Per ragioni di psicotipologia (Kellerman, 1977; Bardel and Lindqvist, 2007), più che la pronuncia /x/ di polacco e tedesco sembra particolarmente insidiosa la pronuncia /ʃ/ dell'inglese e dello spagnolo.

5. METODOLOGIA

Lo studio presenta i risultati di 32 studenti polacchi principianti, iscritti al primo anno del corso di studi in italianistica di un'università polacca⁵, divisi in un gruppo sperimentale (n=17) e uno di controllo (n=15). Al momento dell'esperimento gli apprendenti avevano preso parte a una sola lezione di lingua dedicata all'ortografia. La conoscenza della lingua bersaglio era quindi estremamente limitata, tanto che pare possibile escludere eventuali strategie di decodifica basate sul riconoscimento olistico delle parole bersaglio, il quale richiede notevole pratica ed esposizione all'*input*.

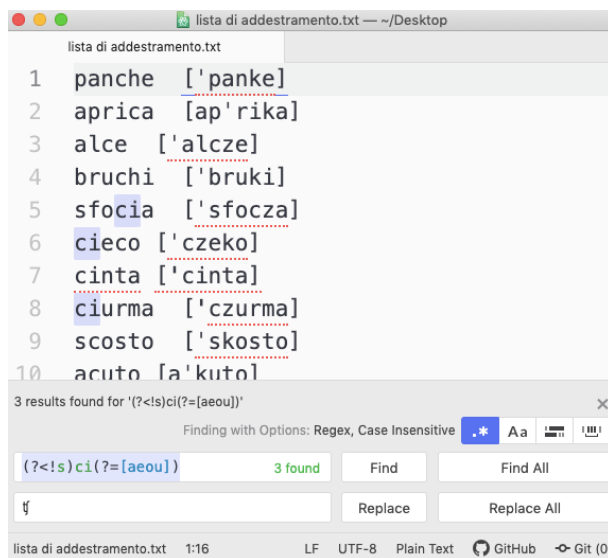
Ai partecipanti del gruppo sperimentale è stato chiesto di portare a lezione i propri pc, sui quali era stato preventivamente installato l'editor di testo *Atom*⁶. Dapprima si è illustrato brevemente il funzionamento delle espressioni regolari con l'aiuto degli esempi polacchi riportati nella sezione 2. I partecipanti hanno poi ricevuto per email una lista di parole italiane selezionate in modo da esemplificare le regole di pronuncia della lettera <c>. Di ciascun elemento la lista indicava la rappresentazione ortografica e la trascrizione fonologica secondo le regole ortografiche del polacco, es. *ciaspole* /'czaspole/. Le parole di esempio sono state selezionate in modo che tutti i suoni fossero rappresentati dai medesimi grafemi tanto in italiano, quanto in polacco, es. /p/, <p>, con la sola eccezione dei grafemi comprendenti <c>. Si sono inoltre evitati i suoni italiani non previsti dalla fonologia del polacco.

La consegna dell'esercizio chiedeva di identificare l'insieme di stringhe di ricerca necessarie per individuare i gruppi formati dalla lettera <c>, al fine di sostituirli con la corrispondente trascrizione fonologica. A questo scopo, i partecipanti potevano elaborare le espressioni regolari nel modo che ritenevano più opportuno (anche con carta e penna), per poi controllarne l'efficacia utilizzando la funzione "trova" del programma "Atom" (Figura 1). Tale metodologia permette di verificare immediatamente se un'espressione regolare è adeguata per identificare tutti e soli i contesti ortografici corrispondenti a un dato fonema.

⁵ Alcuni altri studenti hanno dichiarato di aver studiato italiano anche prima dell'inizio del corso: questi hanno preso parte all'esperimento, ma i loro risultati non saranno considerati nella trattazione.

⁶ <https://atom.io>.

Figura 1. Elaborazione delle espressioni regolari con il programma Atom



Alla fase di addestramento è stata dedicata un'ora di tempo. Al momento della raccolta dati è stata distribuita una nuova lista di parole italiane, di cui si chiedeva di fornire in venti minuti la trascrizione fonologica secondo le convenzioni ortografiche del polacco (tralasciando la posizione dell'accento). Il gruppo di controllo ha partecipato solo a quest'ultima fase, mentre il resto della lezione è stata dedicata a un'altra attività.

Le parole utilizzate nell'esperimento sono state selezionate in modo da risultare poco frequenti e perciò presumibilmente ignote ai partecipanti, avendo particolare cura di escludere quegli elementi per i quali esiste in polacco una parola etimologicamente imparentata. Esse sono dunque equivalenti alle non-parole spesso usate nei protocolli sperimentali per verificare l'assenza di elaborazione semantica nella lettura.

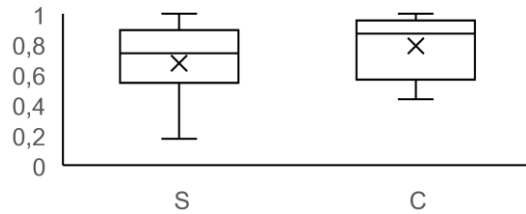
6. RISULTATI

Prima di fornire una descrizione quantitativa dei dati è opportuno presentare una tipologia degli errori riscontrabili nella trascrizioni prodotte dagli apprendenti, così da contestualizzare adeguatamente la successiva analisi. Negli esempi al punto 2 le trascrizioni prodotte dagli apprendenti secondo l'ortografia polacca sono reinterpretate in alfabeto fonetico IPA. Gli errori rilevati riguardano tutti i grafemi considerati, ai quali spesso è associata più di una interpretazione fonologica errata.

- (2) a. *promiscuo* > /promifuo/, /promistfuo/
- b. *rovesciate* > /roveskjate/, /rovesʃate/
- c. *shedammo* > /sedammo/, /stʃedammo/
- d. *vicinato* > /vikinato/
- e. *cedettero* > /ʃedettero/

L'analisi visuale dei risultati (Grafico 1) non sembra evidenziare differenze eclatanti nel punteggio raggiunto dai due gruppi, che appare anzi leggermente più alto e meno disperso nel caso del gruppo di controllo.

Grafico 1. *Punteggi medi del gruppo sperimentale (S) e del gruppo di controllo (C)*



Utilizzando il pacchetto lme4 (Bates et al., 2015) per R (R Core team, 2017), si è proceduto a verificare statisticamente le osservazioni riportate poc'anzi, impostando a questo scopo un modello lineare generalizzato misto (GLMM, Baayen, 2008) con struttura dell'errore binomiale. La variabile dipendente, di tipo binario, è data dalla correttezza della trascrizione del grafema bersaglio. Le variabili indipendenti comprendono l'appartenenza del partecipante al gruppo sperimentale o a quello di controllo, il tipo di bersaglio (secondo la Tabella 3) e la loro interazione. Gli effetti casuali sono rappresentati dall'intercetta per ciascun partecipante e il coefficiente per ciascun tipo di elemento bersaglio. Il risultato del modello è proposto nella Tabella 5, elaborata con il pacchetto sjPlot (Lüdtke, 2018).

Tabella 5. *Esito del modello statistico*

Predictors	Odds Ratios	CI	p
(Intercept)	41.34	8.32 – 205.34	<0.001
gruppo [S]	0.41	0.05 – 3.37	0.407
tipo [1b]	0.04	0.01 – 0.18	<0.001
tipo [1c]	6.33	0.61 – 65.81	0.123
tipo [2]	0.46	0.09 – 2.23	0.334
tipo [3]	0.05	0.01 – 0.23	<0.001
gruppo [S] * tipo [1b]	1.09	0.17 – 6.88	0.924
gruppo [S] * tipo [1c]	0.25	0.01 – 4.51	0.346
gruppo [S] * tipo [2]	1.13	0.15 – 8.41	0.905
gruppo [S] * tipo [3]	0.80	0.11 – 5.94	0.825
Random Effects			
σ^2	3.29		
τ_{00} part	4.21		
τ_{11} part.tipo1b	0.13		
τ_{11} part.tipo1c	1.56		
τ_{11} part.tipo2	1.39		
τ_{11} part.tipo3	1.33		
ρ_{01}	0.15		
0.69			
-0.89			
0.68			
ICC	0.62		
N part	32		
Observations	736		
Marginal R2 / Conditional R2	0.251 / 0.715		

La variabile relativa alla principale domanda di ricerca, cioè l'appartenenza al gruppo sperimentale o a quello di controllo, non risulta significativa. L'interazione tra gruppo e tipo di bersaglio, esplorata mediante il pacchetto *emmeans* (Lenth, 2020), non ha rilevato contrasti significativi. Per quanto riguarda invece l'effetto del tipo di bersaglio, solo i punteggi relativi ai nessi <che, chi> (anche preceduti da <s>) e <sci> risultano significativamente inferiori a tutti gli altri.

7. DISCUSSIONE

Sulla base dei risultati ottenuti, l'esercizio metalinguistico non sembra aver sortito un effetto positivo sulla competenza ortografica dei partecipanti. In termini quantitativi, i punteggi del gruppo sperimentale non differiscono in misura statisticamente significativa da quelli del gruppo di controllo; dal punto di vista qualitativo, gli errori identificati risultano coerenti con quelli riportati dagli studi discussi nella sezione 1.1. Si rilevano differenze statisticamente significative nei punteggi ottenuti per specifiche strutture bersaglio, ma la mancanza di un'interazione con la variabile "gruppo" suggerisce che l'esercizio non abbia influito in modo particolare su specifiche strutture bersaglio e non altre.

Si potrebbe quindi concludere che l'attività proposta sia di scarsa efficacia ai fini dello sviluppo delle abilità ortografiche, almeno per quanto riguarda il campione considerato. Quest'ultima precisazione non è un dettaglio di poco conto: come detto, si tratta di studenti iscritti al primo anno di laurea triennale, esaminati nei primissimi giorni di studio dell'italiano. Dal momento che la riflessione linguistica esplicita non è sempre praticata nella scuola secondaria, è possibile che la competenza metalinguistica di questi studenti principianti fosse ancora immatura al momento della raccolta dati. L'esercizio proposto potrebbe dunque essersi rivelato piuttosto oscuro sia nello svolgimento, sia anche negli obiettivi.

Vi è poi la difficoltà legata alla sintassi delle espressioni regolari stesse. Nonostante se ne siano presentate sole tre, la natura altamente astratta degli operatori le rende poco intuitive. È una carenza metodologica dello studio il fatto che la competenza dei partecipanti nell'uso di questi strumenti non sia stata in alcun modo verificata: dato l'ambito di applicazione piuttosto tecnico delle espressioni regolari, tuttavia, si può supporre che il loro funzionamento fosse ignoto alla maggior parte degli studenti. A tutto ciò si aggiunge che la fase di addestramento è durata complessivamente un'ora, forse poco alla luce della complessità della struttura bersaglio, della novità dello strumento e dell'inesperienza dei partecipanti.

In conclusione, sembrerebbe che l'esercizio sia stato proposto in condizioni sfavorevoli, o piuttosto, che se ne siano fraintesi l'obiettivo e la tempistica. La sua efficacia andrebbe dunque verificata nuovamente in differenti condizioni sperimentali. La variabile cruciale a questo proposito pare essere la scelta dei partecipanti. In particolare, sarebbe auspicabile una maggiore confidenza con alcuni strumenti informatici, quali in particolare le istruzioni a riga di comando se non le stesse espressioni regolari. Nell'ambito dell'approccio qui proposto, in effetti, queste non rappresentano un fine bensì soltanto un mezzo, così che la loro conoscenza preventiva potrebbe permettere ai partecipanti di concentrarsi meglio sulla logica dell'esercizio. Sarebbe senz'altro opportuna anche una maggiore consapevolezza metalinguistica, per quanto il suo sviluppo rimanga uno degli obiettivi intrinseci dell'esercizio. Alla luce di queste considerazioni, i candidati più naturali sembrano gli studenti dei numerosi corsi in umanistica digitale. Oltre che all'inizio di un corso di lingua, l'attività si potrebbe proporre anche nell'ambito di un corso di linguistica, in particolare dopo aver introdotto i concetti fondamentali della priorità del parlato sullo

scritto e aver svolto abbondante pratica di trascrizione fonetica. In queste condizioni, l'attività potrebbe offrire la possibilità di applicare le competenze metalinguistiche dei partecipanti a un nuovo strumento dalle notevoli potenzialità (si pensi alla trascrizione automatica), consolidando nel contempo le competenze ortografiche degli studenti nella lingua straniera da essi studiata (nel caso in cui l'esercizio sia applicato a quest'ultima). Che tale occasione di pratica non sia superflua è dimostrato dalla rassegna presentata nella sezione 1.1, la quale mostra che diverse incertezze sulla grafia di taluni fonemi permangono anche in stadi piuttosto avanzati di competenza. In quanto esercizio metalinguistico astratto, d'altra parte, il lavoro con le espressioni regolari può anche essere applicato alla L1 degli studenti o a qualunque altra lingua.

Riassumendo, l'analisi dell'ortografia di una L2 mediante le espressioni regolari non sembra essersi rivelata una metodologia glottodidattica efficace. La ricerca futura potrebbe però indagare l'utilità di questo approccio alla luce di obiettivi didattici diversi, quali lo sviluppo della competenza metalinguistica e di abilità informatiche utili nell'ambito dell'umanistica digitale.

8. CONCLUSIONI

Con il presente studio si è voluto valutare l'efficacia di un'attività metalinguistica volta a potenziare la competenza ortografica di un gruppo di apprendenti polacchi di italiano L2. In particolare, si chiedeva di formalizzare le regole di lettura della lettera <c> utilizzando la sintassi altamente astratta delle espressioni regolari, comunemente in uso in informatica. L'attività non sembra avere sortito l'effetto atteso, almeno per quanto riguarda l'elaborazione delle espressioni regolari richieste. La riflessione linguistica in cui gli studenti necessariamente si sono impegnati, al contrario, potrebbe aver sortito effetti favorevoli sul lungo periodo, non misurabili però nell'ambito di questo studio.

La ragione di tale fallimento è da individuarsi principalmente nel pubblico a cui è stato proposto l'esercizio: trattandosi di studenti principianti, con una competenza metalinguistica ancora immatura e probabilmente poco esperti di strumenti informatici, appare verosimile che siano sorte serie difficoltà sia nell'individuazione della logica del compito, sia nella sua realizzazione tecnica. Lo studio conclude perciò che se pure l'attività proposta è poco adatta allo sviluppo della competenza ortografica in ambito strettamente glottodidattico, tuttavia se ne potrebbe indagare l'utilità per altri scopi, quali il potenziamento della competenza metalinguistica di studenti più maturi, oppure ancora la familiarizzazione con le applicazioni informatiche degli strumenti teorici del linguista.

RIFERIMENTI BIBLIOGRAFICI

- Albano Leoni F. (2009), *Dei Suoni e Dei Sensi: Il Volto Fonico Delle Parole*, il Mulino, Bologna.
- Angeles P., Notarnicola A., Costabile D., Marinelli C., Judica A., Zoccolotti P., Luzzatti C. (2008), *DDO-2: diagnosi dei disturbi ortografici in età evolutiva*, Erickson, Trento.
- Baayen H. (2008), *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*, Cambridge University Press, Cambridge.
- Bardel C., Lindqvist C. (2007), "The role of proficiency and psychotypology in lexical cross-linguistic influence. A study of a multilingual learner of Italian L3", in Chini M., Desideri P., Favilla M. E., Pallotti G. (a cura di), *Atti del VI Congresso*

- Internazionale dell' Associazione Italiana di Linguistica Applicata*, Guerra, Perugia, pp. 123-145.
- Bates D., Maechler M., Bolker B., Walker S. (2015), "Fitting Linear Mixed-Effects Models Using lme4", in *Journal of Statistical Software*, 67, pp. 1-48:
<https://www.jstatsoft.org/article/view/v067i01>.
- Bažec H. (2019), "Pronuncia e grafia tra L1 e L2: alcune considerazioni sugli errori ortografici di studenti slovenofoni", in *Italica Wratislaviensia*, 10, pp. 11-29.
- Best C. (1995), "A direct realist view of cross-language speech perception: New Directions in Research and Theory", in Strange W. (Ed.), *Speech perception and linguistic experience: issues in cross-language research*, York Press, Baltimore, pp. 171-204.
- Bhide A. (2015), "Early literacy experiences constrain L1 and L2 reading procedures", in *Frontiers in Psychology*, 6, pp. 1-16 :
<https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01446/full>.
- Bozzo M. T., Pesenti E., Siri S., Usai M. C., Zanobini M. (2000), *Test CEO: classificazione degli errori ortografici*, Erickson, Trento.
- Corino E., Marengo C. (2009), "Elicitare scritti a partire da storie disegnate: il corpus di apprendenti VALICO", in Andorno C., Rastelli S. (a cura di), *Corpora di italiano L2: tecnologie, metodi, spunti teorici*, Guerra, Perugia, pp. 113-128.
- Cossu G., Shankweiler D., Liberman I. Y., Katz L., Tola G. (1988), "Awareness of phonological segments and reading ability in Italian children", in *Applied Psycholinguistics*, 9, pp. 1-16.
- van Daal V. H. P., Wass M. (2017), "First- and Second-Language Learnability Explained by Orthographic Depth and Orthographic Learning: A "Natural" Scandinavian Experiment", in *Scientific Studies of Reading*, 21, pp. 46-59.
- Ehri L. C. (2005), "Learning to Read Words: Theory, Findings, and Issues", in *Scientific Studies of Reading*, 9, pp. 167-188.
- Ellis N. C., Natsume M., Stavropoulou K., Hoxhallari L., Daal V. H. P., Polyzoe N., Tsipa M.-L., Petalas M. (2004), "The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts", in *Reading Research Quarterly*, 39, pp. 438-468.
- Flege J. (1995), "Second-language speech learning: Theory, findings, and problems", in Strange W. (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, York Press, Timonium, MD, pp. 229-273.
- Frith U. (1986), "A developmental framework for developmental dyslexia", in *Annals of Dyslexia*, 36, pp. 67-81.
- Gataveckaitė A. (2013), *L'italiano di apprendenti lituani: elementi acquisizionali e implicazioni didattiche*, Tesi di dottorato, Università di Bologna.
- Katz L., Frost R. (1992), "The Reading Process is Different for Different Orthographies: The Orthographic Depth Hypothesis", in Frost R., Katz L. (Eds.), *Advances in Psychology*, Elsevier, Amsterdam, pp. 67-84.
- Kellerman E. (1977), "Towards a characterization of the strategy of transfer in second language learning", in *Interlanguage Studies Bulletin*, 2, pp. 58-145.
- Lenth R. (2020), *Emmeans: Estimated Marginal Means, Aka Least-Squares Means*:
<https://www.rdocumentation.org/packages/emmeans/versions/1.5.0>.
- Lüdecke D. (2018), *SjPlot - Data Visualization for Statistics in Social Science*, Zenodo:
<https://zenodo.org/record/2400856/export/hx#.X1-I1NRS-pp>.
- R Core team (2017), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna.
- Seymour P. H. K., Aro M., Erskine J. M. (2003), "Foundation literacy acquisition in European orthographies", in *British Journal of Psychology*, 94, pp. 143-174.

Shoemaker E. (2015), “The Development of Perceptual Sensitivity to Polish Sibilants at First Exposure”, in *Proceedings of the Annual Meetings on Phonology 2*: <https://journals.linguisticsociety.org/proceedings/index.php/amphonology/article/view/3770/3461>.

Tressoldi P., Cornoldi C., Re A.M. (2013), *Batteria per la valutazione della Scrittura e della competenza ortografica-2*, Giunti, Firenze.