Alsic

Apprentissage des Langues et Systèmes d'Information et de Communication

Vol. 26, n° 3 | 2023
Varia
Pratique et recherche

# ICT and metalinguistic competence: Developing phonetic transcription skills through regular expressions

Jacopo Saturno
https://doi.org/10.4000/alsic.6554

## *Résumés*

Français English
L'article décrit une expérience pédagogique visant à évaluer la pertinence des expressions régulières (regex), un outil typique des TIC (technologies de l'information et de la communication), pour améliorer la capacité d'un groupe d'étudiants inscrits en langue étrangère à identifier la relation entre les unités graphiques et phoniques de leur L1, l'italien. Dans cette langue, plusieurs sons sont systématiquement représentés par des graphèmes complexes comprenant plus d'une lettre, par exemple <sci> = /ʃ/. De plus, la prononciation de tels graphèmes varie en fonction du contexte graphique : par exemple, la lettre <i> fonctionne exclusivement comme un signe diacritique si elle est suivie d'une voyelle, alors qu'elle encode également le son /i/ si elle est suivie d'un autre son ou si elle se trouve en position finale du mot (voir scimmia /'ʃim:ja/ "singe" vs sciame /'ʃame/ "essaim"). De telles séquences complexes peuvent être exprimées à l'aide d'expressions régulières reconnaissant le contexte graphique en termes abstraits, y compris des spécifications telles que "en position finale du mot" ou "suivi d'une lettre autre que <a>, <o>, <u>". Dans l'expérience décrite dans cet article, 255 étudiants italiens sans expérience d'utilisation de regex (expressions régulières) ont été invités à développer un ensemble d'expressions régulières reflétant les règles qui régissent la transcription de la lettre <c> en italien, dont la prononciation varie entre /k/, /ʧ/ et /ʃ/ selon le contexte graphique. Le but de l'expérience était de vérifier si la réflexion métalinguistique approfondie requise par l'exécution de cette tâche peut affecter positivement les compétences de transcription des participants, en partant du principe que si la syntaxe technique et les opérateurs des regex sont suffisamment maîtrisés, ils peuvent être exploités pour produire inductivement une description métalinguistique d'un phénomène linguistique. Afin de documenter les compétences initiales des participants avant la tâche expérimentale, ils ont d'abord été invités à décrire verbalement les règles régissant la prononciation de la lettre <c> en italien et à transcrire phonologiquement une liste de mots qui contiennent cette lettre. Ensuite, les participants ont dû formaliser les règles de prononciation de la lettre <c> à travers un ensemble approprié de regex, en appliquant l'outil de recherche et remplacement d'un éditeur de texte à un ensemble de mots cible. Enfin, les

participants ont de nouveau été invités à décrire les règles de prononciation de <c> et à transcrire phonologiquement un nouvel ensemble de mots cible afin de vérifier si la tâche expérimentale améliorait leurs compétences métalinguistiques et leurs capacités de transcription. Les résultats montrent une différence évidente dans les scores de toutes les tâches selon que la lettre <i> fonctionne comme un signe diacritique ou non. De plus, à la suite de la tâche expérimentale, le nombre d'erreurs de transcription affectant les mots contenant le signe diacritique <i> a diminué de 29% à 16%. Plus important encore, les descriptions de l'orthographe italienne des participants se sont également améliorées au fil du temps. La réflexion nécessaire pour formaliser un ensemble approprié de regex a donc eu un impact positif sur la précision de transcription et les capacités de description métalinguistique des étudiants. On peut conclure que l'étude des expressions régulières, dont les principales applications sont les technologies de l'information et de la communication, peut être appliquée de façon pertinente au développement des compétences de réflexion métalinguistique, qui, à leur tour, affectent positivement les compétences typiquement linguistiques telles que la transcription phonétique.
The present paper describes an applied linguistics experiment aiming to evaluate the appropriateness of regular expressions (regex) to improve foreign language students' ability to identify the relationship between the graphic (letters) and phonic (phonemes) units of a language. Since the focus of the study is on metalinguistic rather than linguistic skills, the native language of the participants–L1 Italian–was chosen to this purpose. Specifically, regexes were used to formalise the phonological transcription rule of a set of graphemes (eg <scia> /ʃa/) well known for causing transcription errors due to the interference of orthography. Both before (T1) and after (T2), the regex formalisation task, 255 university students of foreign languages were asked to verbally explain the relevant transcription rules and transcribe a set of words containing the grapheme of interest. The results show that indeed, the accuracy of transcriptions and metalinguistics accounts significantly improved after the task. A discussion of the possible applications of regexes in the training of students of foreign languages and linguistics concludes the paper.

### *Entrées d'index*

**Mots-clés :** métalinguistique, regex, orthographe, italien, langue maternelle
**Keywords:** language awareness, regex, spelling, Italian, mother tongue
**Rubriques :** Pratique et recherche

### *Texte intégral*

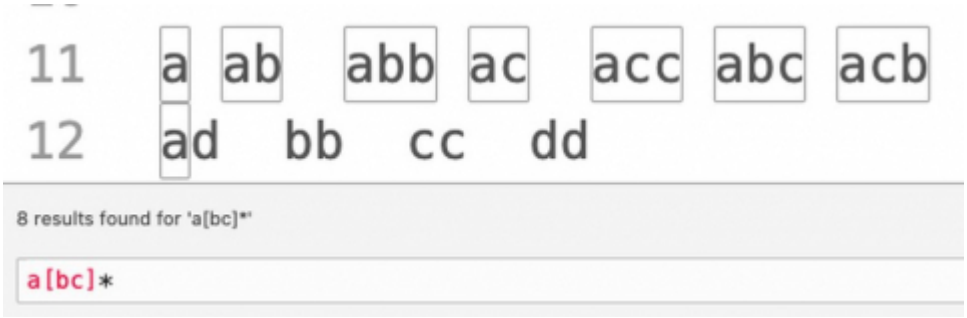
# 1. Introduction

1 The present study intends to assess the suitability of regular expressions (henceforth "regex") for the development of metalinguistic competence among students of foreign languages. A regex is a string of characters that can be interpreted by a regex engine in order to locate matching patterns in a text. Crucially, for the purposes of the present study (see below), a regex is not limited to exact matches, but can encode abstract conditions to be matched in the input text, such as for instance "any character <x> not preceded by character <y>." Regexes are typically often used as a technical tool in computer science and computational linguistics but are decidedly less common in the training toolkit of students of foreign languages. Coherently, the participants in the experiment reported here only had basic ICT (information and communication technology) skills, in no case comprising regexes or related tools, such as programming languages, command-line input, etc. They were introduced to regexes in the framework of a course in "ICT and language data", aiming to provide participants with basic technical skills needed to process linguistic data (eg transcription of aural data, manipulation of linguistic strings, computing of descriptive statistics, etc.). Within this course, an experiment was run to investigate whether the formalisation of a phonetic transcription system using regexes positively affects the students' ability to identify the relationship between the graphic (letters and groups of letters) and phonic (sounds) units of their native language, Italian.

2 The task may be interpreted as the elaboration of a semi-automatic system of phonetic transcription based on the following rationale: Suppose one wants to

phonologically transcribe the word *scoglio* /'skoʌo/ "rock" using the "find and replace" command of a text editor. In order to do that, one needs to identify the orthographic representation of each phoneme and substitute it with the corresponding symbol in a phonetic alphabet such as the IPA[1] (Landau et al., 1999). In Italian, some of the letters that compose orthographic words do not correspond to a single sound (and, therefore, to a single character of a phonetic alphabet), since their pronunciation varies depending on the graphic context (the surrounding letters). To exemplify, the letter <l> corresponds to the sound /l/ unless it is preceded by <g> and followed by <i>, in which case the three letters form the trigraph <gli>, pronounced /ʎ/. The rule, however, is more complex than that, since within these complex graphemes (composed of more than one letter), the status of the letter <i> again depends on the graphic context. If it is followed by a vowel, <i> functions exclusively as a diacritic and is not represented in a phonetic transcription, eg *scoglio* /'skoʌo/ 'rock'. If, on the other hand, it is followed by another sound (eg *inorgoglire* /inorgoʌire/ 'become proud'), or occurs in word-final position (eg *scogli* /'skoʌi/ "rocks"), the letter <i> also corresponds to the sound /i/. In order to adequately transcribe words containing the trigraph <gli> through a "find-and-replace" command, therefore, one cannot just specify a fixed search pattern such as <gli>, which would incorrectly transcribe *scogli* as /'skoʌ/; instead, one needs to specify a string acknowledging the graphic context in abstract terms, including such specifications as "word-final position", or "followed by a letter other than <a>, <o>, <u>." Such complex patterns can be effectively expressed using regexes (Goyvaerts & Levithan, 2009). To exemplify, the single regex <a[bc]*> identifies all occurrences of <a> followed by a variable number (including 0) of <b> or <c>, example <a, ab, ac, abb, acc> (Figure 1). The pattern is "abstract" in the sense that one does not need to specify the concrete results of the search (the specific strings identified by the regular expression), but only the requirements that they need to satisfy. Regexes can thus radically reduce the number of "find-and-replace" operations required to phonetically transcribe a text.

**Figure 1–Using regexes to find abstract patterns in the text editor Atom.**




3 In the present study, a group of students of foreign languages were asked to develop a set of regexes reflecting the rules that govern the transcription of the letter <c> in Italian. The aim of the experiment was to verify whether the in-depth metalinguistic reflection required by this task could positively affect the participants' phonological transcription skills, due in turn to an improvement in their awareness of the Italian system of phoneme-grapheme correspondences. The rationale is that if the technical syntax and the operators of regexes are mastered sufficiently, they can be exploited to inductively produce a metalinguistic description of a linguistic phenomenon.

4 The orthography of Italian is particularly well suited for this purpose. Since the relationship between graphical and phonic units is almost univocal, it can be classified as quite transparent (Katz & Frost, 1992). At the same time, however, several sounds are systematically represented by complex graphemes comprising more than one letter, eg <sci> = /ʃ/ in *uscire* /u'ʃire/ "to leave", which partially obscures the relation between orthography and pronunciation. Additionally, the relation between phones and graphemes is not biunivocal in that the same sound may be represented by more than one grapheme (eg /k/ = <c> and <ch> in *chimica* /'kimika/ "chemistry"); conversely, the same grapheme may correspond to different sounds depending on the orthographic

context (eg <c> = /k/ in *cane* /'kane/ 'dog', /ʃ/ in *scena* /'ʃena/ 'scene', /ʧ/ in *cena* /'ʧena/).

5 Indeed, some points of Italian orthography seem particularly problematic for learners of L2 Italian (Guerrero, 2016; Saturno, 2020); the utility of phonetic transcription, on the other hand, is a matter of discussion in the domain of both language teaching and teacher training (Torresan, 2010; Calabrò, 2015). In the literature on pupils' and students' competences in L1 Italian (excluding the case of students with language-specific impairments), orthographic errors are rarely mentioned, most probably due to the attention they are paid in schooling: indeed, a lack of mastery of this language layer, though not necessarily detrimental to effective communication, certainly constitutes a major social stigma (Berruto 2012, p. 48-49). If present at all, examples of orthographic difficulties in the written production of L1 Italian speakers mainly concern diacritics (eg <è>, <é>, <'>, etc.) or the special cases in which orthography deviates from the phonological principle for etymological or customary reasons (Viale, 2011, p. 148). At the same time, these studies highlight significant metalinguistic gaps, concerning amongst others the notion of "subject" (Calaresu & Dal Negro, 2018) or the rationale behind the choice of verb forms (Andorno et al. 2018, p. 39). On the other hand, De Masi and Maggio (2006) note that among university students, some of the greatest difficulties are recorded in the field of phonetics and phonology. Concerning phonetic transcription–a tool that is only taught in university-level courses related to linguistics–, Lavinio (2011) reports that even students enrolled in a master course (MA) in Educational Linguistics–in principle language specialists–suffered from a marked tendency to insert the sound /i/ in the phonological transcription of graphemes in which the letter <i> functions exclusively as a diacritic, eg *figlio* */'fiʌio/, cf. /'fiʌio/ "son." Moreover, judging by the students' definitions of a set of technical notions, it is also clear that the reflection on phonetics and phonology very often starts from the orthographic representation of language (to exemplify, the phoneme is defined as "the sound of a single *letter*", p. 285). Against this picture, it must be noted that duly acknowledging these difficulties, most Italian introductory textbooks in linguistics do place great emphasis on phonetic transcription in their exercise section, with a particular focus on the diacritic role of <i>. Typical tasks include the transcription of words comprising problematic graphemes, eg *sciogliere* /'ʃoʌere/ "melt" (Berruto & Cerruti, 2017, p. 87), and the identification of minimal pairs potentially masked by presence of a diacritic in one of the terms, eg *piglia* /'piʌa/ 'take:3SG' vs *pigna* /'piɲa/ "pinecone" (Berruto & Cerruti, 2017, p. 89).

# 2. The target structure and its formalisation through regexes

6 The present study focuses on the letter <c>, whose pronunciation varies between /k/, /ʧ/ and /ʃ/ depending on the graphic context (Table 1). Again, the letter <i> deserves particular attention.

**Table 1–Rules governing the pronunciation of clusters containing <c> in Italian.**

| IPA | Preceding context | Following context | example |
|---|---|---|---|
| /k/ | any | <h>; ≠ <i, e> | *cane, chilo, credo* |
| /ʧ/ | ≠ s | <e>; <i> followed by <a>, <e>, <o>, <u> | *cena, ciao, cielo, ciondolo, ciuffo* |
| /ʧi/ | ≠ s | <i> not followed by <a>, <e>, <o>, <u> | *cinema, luci* |

| | | | |
|---|---|---|---|
| /ʃ/ | s | <e>; <i> followed by <a>, <e>, <o>, <u> | *scena, sciame, scienza, sciolto, sciupare* |
| /ʃi/ | s | <i> not followed by <a>, <e>, <o>, <u> | *scivolo, rovesci* |

7 These rules can be formalised using the regexes described in table 2 and exemplified using the words in (1).

(1)
*cane* /'kane/ "dog"
*cena* /'ʧena/ "dinner"
*cima* /'ʧima/ "summit"
*cono* /'kono/ "cone"
*curva* /'kurva/ "turn"
*sciame* /'ʃame/ "swarm"
*scena* /'ʃena/ "scene"
*scimmia* /'ʃim:ja/ "monkey"

**Table 2–Regexes discussed in the course.**

| regex | meaning | example | result |
|---|---|---|---|
| [xyz] | Any of the characters comprised in the square brackets. | c[aeo] | *ca* in *cane*, *ce* in *cena* and *scena*, *co* in *cono* |
| [^xyz] | Any character other than those comprised in the square brackets. | c[^aeo] | *ci* in *cima*, *sciame* and *scimmia*, *cu* in *curva* |
| \| | or | ca\|me | *ca* in *cane*, *me* in *sciame* |
| x(?=y) | x followed by y, y not being part of the match. | ci(?=m) | *ci* in *cima* and *scimmia* (but not in *sciame*) |
| x(?!y) | x not followed by y, with y not being part of the match. | ci(?!m) | *ci* in *sciame* (but not in *cima* and *scimmia*) |
| (?<=y)x | x preceded by y, y not being part of the match. | (?<=s)ci | ci in sciame and scimmia (but not cima) |
| (?<!y)x | x not preceded by y, y not being part of the match. | (?<!s)ci | *ci* in *cima* (but not in *sciame* and *scimmia*) |

8 Particular attention should be paid to the operators comprising round brackets (technically *lookahead* and *lookbehind*, depending on whether the brackets follow or precede the matched string, respectively), which indicate that the content of the parenthesis must necessarily be present (operators with <=>) or absent (operators with <!>), without, however, being itself part of the match. To exemplify, the expression "sc(?=e|i[^aou])" will match all occurrences of <sc> followed by <e> or <i>, the latter in turn not followed by <a,o,u>, without, however, matching the vowel. A single find-and-replace operation can thus be performed, turning <sc> into /ʃ/ and leaving the vowel unchanged (Italian vowel graphemes are normally identical to the corresponding IPA symbol). A set of regexes sufficient to transcribe all Italian graphemes containing <i> is presented in (2).

(2)
a. c(h|(?=[aou])) = /k/
b. (?<!s)c((?=e)|i(?=[aeou])|(?=i[^aeou])) = /ʧ/
c. sc((?=e)|i(?=[aeou])|(?=i[^aeou])) = /ʃ/

The same rule can be reformulated in other ways, as exemplified in (3). The two set of expressions are equivalent in terms of efficiency and complexity, although (2) does seem preferable from a linguistic point of view because it highlights the fact that the strings in b. and c. are identical in everything except the letter preceding <c>.

(3)
a. ch|c(?![eih]) = /k/
b. (?<!s)c(?=e|i[^aou])|(?<!s)ci(?=[aou]) = /ʧ/
c. sc(?=e|i[^aou])|sci(?=[aou]) = /ʃ/
The complex expressions in (2) and (3) can also be broken down into a greater number of simpler, more specific (ie less general) regexes. To exemplify, the two expressions in (4) split up the regex presented in (2a) and (3a).
(4)
a. ch = /k/
b. c(?=[aou]) = /k/

9 Unnecessarily from a technical point of view, but perhaps to the benefit of transparency, one may also specify regexes affecting both the letter <c> itself and part of the surrounding context. To exemplify, while the regex in (5a) only modifies the letter <c> and leaves unchanged the existing <e>, (5b) substitutes the whole group <ce>.
(5)
a. (?<!s)c(?=e) = /ʧ/
b. (?<!s)ce = /ʧe/

10 Finally, one may also reduce the complexity of the regex set by properly manipulating the order in which the "find and replace" operations are performed, as in the flow described in table 3. Phase 2 relies on the assumption that all the instances in which <sci> is followed by a vowel have been matched and substituted in phase 1: as a result, only those cases in which <sci> is in word-final position or is followed by a consonant can still be found in the text. This approach is very sensitive to the order of the substitutions: performing phases 3 and 4 before phases 1 and 2 would make it impossible to adequately transcribe the groups <sci>, which by then would appear as <sʧ>.

**Table 3–Sequential find-and-replace.**

| sequence | regex | replacement |
|---|---|---|
| 1 | sci(?=[aeou]) | ʃ |
| 2 | sci | ʃi |
| 3 | ci(?=[aeou]) | ʧ |
| 4 | ci | ʧi |

11 For the purpose of this study, in any case, a sequential workflow still requires careful metalinguistic reflection in order to account for the status of the letter <i> depending on the graphic context. The difference between the preferable solution in (3) and the sub-ideal alternative in (5b) or table 3 concerns the technical details of regex elaboration, rather than metalinguistic competence. For the purpose of the present study, all alternatives presented so far are equally acceptable.

# 3. Research questions

12 Globally, the present paper aims to verify whether the study of regexes can positively affect the metalinguistic competence of students of foreign languages. In particular, the following research questions can be formulated as follows:

13 RQ1: Does the regex-based task positively affect the students' ability to phonetically transcribe grapheme clusters containing the letter <c>, especially when the letter <i> functions as a diacritic?
HYP1: The participants' phonetic transcription skills are expected to benefit from the experimental task, especially concerning the cases in which <c> in combination with

the diacritic <i> represents the sounds /ʃ/ and /ʧ/. A ceiling effect is expected for the cases in which <c> = /k/.

14 RQ2: Does the task positively affect the students' ability to explain the regularities that govern the pronunciation of the letter <c>, especially concerning the diacritic role of the letter <i>?
HYP2: The experimental task is expected to improve the participants' explanations by highlighting the fact that the letter <i> may function as a diacritic, which is likely to otherwise go unnoticed by most participants.

15 RQ3: Is there a correlation between the participants' accuracy of metalinguistic descriptions and their phonetic transcription scores?
HYP3: A strong correlation is expected: the ability to adequately describe pronunciation rules should translate in appropriate transcriptions and vice versa.

16 RQ4: Are the technical difficulties associated with the use of regexes within the reach of students enrolled in a computer science course for students of foreign languages?
HYP4: It is expected that the limited number of operators necessary to formalise the spelling of Italian, together with the time devoted to them during the course, can make the required set of regexes accessible to all students.

17 It should be noted that improvement in the phonological transcription (RQ1) and metalinguistic description (RQ2) of the target structure does not directly depend on the correct identification of regexes, but rather on the underlying metalinguistic reflection. Scores in the above-mentioned tasks could significantly improve even if the learner was not technically able to correctly compile the corresponding regexes.

# 4. Methodology

## 4.1. Procedure

18 The experiment involved 255 Italian-speaking MA students of foreign languages enrolled in a course in "ICT and language data."[2] The course did not cover linguistic topics such as phonetics, phonology, or transcription; however, 238 participants declared that throughout their career (bachelor and master), they had taken at least one course discussing phonetic transcription. None declared any knowledge of the particular regex flavour adopted in this study (Perl), although a few did report using basic operators (such as <.> or <?>) to refine searches on the internet or within corpora.

19 The experiment was first run in 2020 and then replicated in 2021 with identical methodology. The module of the course in "ICT and language data" concerning regexes also remained unchanged. Data collection took place through an online form. To start, participants were asked to verbally describe the rules governing the pronunciation of the letter <c> in Italian (phase 1) and to manually type the phonological transcription of a list of words containing it (phase 2) using the ascii-based SAMPA phonetic alphabet (Wells, 1997). SAMPA (Speech Assessment Methods Phonetic Alphabet) was chosen to facilitate the participants' task by removing the potential difficulty associated with the entering of IPA symbols not present on computer keyboards. For reasons of readability, however, within the present paper, transcriptions will be presented in IPA. In order to reduce the complexity of the task, stress marks and other diacritics (such as <:>, indicating consonant length) were to be avoided. Phase 1 and phase 2 had the purpose to document the participants' initial skills prior to the experimental task.

20 In phase 3, the participants received a text file divided into two columns. The former comprised the orthographic representation of a set of 22 Italian words containing the letter <c>; the second presented their SAMPA phonetic transcription. Participants were asked to use this file to formalise the pronunciation rules of the letter <c> by elaborating an appropriate set of regexes. Using the find-and-replace tool of any text

editor supporting regexes, they could experiment different solutions until all entries in column 1 matched their counterparts in column 2 (Figure 2). The present experiment relied on various versions of Atom, now discontinued (Atom, 2020).

**Figure 2–Elaborating regexes through find-and-replace in the text file in Atom.**


21 After this phase, students were again asked to describe the pronunciation rules of <c> (phase 4) and to phonologically transcribe a new set of words (phase 5). These last phases had the purpose to verify whether the experimental task improved the participants' metalinguistic competence and transcription skills. The two occasions when participants were asked to describe Italian orthography and provide a phonological transcription will be referred to as T1 (phases 1 and 2) and T2 (phases 4 and 5), respectively. The structure of the experiment is summarised in table 4.

**Table 4–Structure of the experiment.**

| time | phase | task |
|---|---|---|
| Time 1 | 1 | Enunciation of pronunciation rules |
| | 2 | Phonetic transcription of target items (set 1) |
| Experimental task | 3 | Regex formalization |
| Time 2 | 4 | Enunciation of pronunciation rules |
| | 5 | Phonetic transcription of target items (set 2) |

22 The participants were provided with a table showing the SAMPA symbols required to carry out the task alongside their IPA counterparts, the graphemes representing them in Italian and their articulatory classification.

## 4.2. Target structure

23 At each data collection time, participants were tested on 22 Italian words, which together represents all possible combinations of the sounds represented by the letter

<c> (/k/, /ʃ/, /ʧ/) and the five vowels of Italian (/a/, /e/, /i/, /o/, /u/), with the addition of /r/ in the case of /k/. Although the velar (/k/) pronunciation of <c> is not affected by the presence of a preceding <s> (eg <ca> /ka/, cf. <sca> /ska/), this combination was introduced to highlight the effect of such modification on the groups <ce> and <ci>, eg <cia> /ʧa/, cf. <scia> /ʃa/.

24 Since the participants were provided with a SAMPA reference chart, it seems that as far as the elaboration of regexes is concerned, the only true source of potential difficulty is represented by the letter <i> in the groups <sci> and <ci>. In light of this, four different scores were computed for the transcription and regex formulation tasks (Table 5).

**Table 5–Target items.**

| score | Sound | orthography |
|---|---|---|
| Total | | |
| Distractors | /k/ | <ca>, <che>, <chi>, <co>, <cu>, <cr> |
| | /sk/ | <sca>, <sche>, <schi>, <sco>, <scu>, <scr> |
| | /ʃe/ | sce[3] |
| | /ʧe/ | <ce> not preceded by <s> |
| Target items | /ʃa/ | <scia> |
| | /ʃo/ | <scio> |
| | /ʃu/ | <sciu> |
| | /ʧa/ | <cia> not preceded by <s> |
| | /ʧo/ | <cio> not preceded by <s> |
| | /ʧu/ | <ciu> not preceded by <s> |
| <i> items | /ʃi/ | <ci> not preceded by <s> and not followed by <a>, <o>, <u> |
| | /ʧi/ | <sci> not followed by < a>, <o>, <u> |

25 The rationale for distinguishing target items and <i> items is the following. The substitutions "sci = /ʃi/" or "sc(?=i) = /ʃ/" will produce an erroneous transcription when <i> functions as a diacritic (6a). In contrast, "sci(?=[aeou]) = /ʃ/" will correctly transcribe the groups <scia, scie, scio, sciu>, but will fail to identify the cases in which <i> is followed by another letter (6b).

(6)

a. */'ʃiabola/, cf. target *sciabola* /'ʃabola/ "saber"

b. */'ʃmmja/, cf. target *scimmia* /'ʃimmja/ 'monkey'

26 To identify both types of strings, one needs to either perform two separate "find-and-replace" operations ("sci(?=[aou]) = ʃ" and "sci(?![aou]) = ʃi"), or (better) perform a single search using a regex allowing for an alternative ("sc(?=i[^aou])|sci(?=[aou]) = /ʃ/"). Elaborating two separate scores makes it possible to better characterise the participants' errors (ie regexes leading to erroneous transcriptions like 6a or 6b, respectively). Following the exclusion of non-pertinent responses, the definitive dataset comprises 10671 data points produced by 244 participants, divided into 8047 distractors and 2923 target items.

27 The online form also asked participants to evaluate several aspects of the task on a five-point Likert scale, including the difficulty of three of its components (verbal description of the pronunciation rules, regex formalization, phonetic transcription), the perceived utility of regexes (for the study of foreign languages, for the study of

linguistics, for phonetic transcription), and their overall enjoyment of the activities based on regexes. All data were elaborated and analysed in R (R Core team, 2017) with the aid of a variety of packages (Chang, 2014; Dragulescu & Cole, 2020; McElreath, 2020; Stan Development Team, 2019, 2020; Wickham et al., 2019).

## 4.3. Assessment

28 The set of regexes was first automatically checked for obvious typos or mistakes, such as whitespaces separating alternative terms (7a), missing brackets (7b) or unnecessary elements (<,> in 7c, <|> in 7d).

(7)

a. ch(?=[ei]) | c(?![ei])

b. c =?=[aeiou]

c. c (?=[i,e])

d. sci(?=[a|o|u])|sc(?=[e|i])

29 For each participant, the adequacy of the set of "find-and-replace" operations was then evaluated by applying it–in the order in which the expressions were inserted in the online form–to a list of Italian words. Following the application of the "find-and-replace" operations, the transformed list was compared to its target SAMPA transcription. For each target type (Table 5), a score was computed based on the proportion of correctly transcribed items. Concerning phonetic transcriptions, the only relevant criterion for the purpose of the present study is whether or not they erroneously include <i> or other related symbols, eg <j> in (8).

(8)

*/mantʃia/ or */mantʃja/, cf. target *mancia* /'mantʃa/ "tip"

30 As long as they did not interfere with the criterion stated above, obvious typing mistakes and inaccuracies due to the participant's scarce familiarity with SAMPA were ignored. Finally, metalinguistic descriptions were manually assessed by the author of the paper. Again, the only relevant criterion was whether or not the participants acknowledged the diacritic function of the letter <i> in the relevant graphemes. The metalinguistic analysis was considered correct even if the diacritic role of <i> was acknowledged with respect to only one of the two graphemes (<ci, sci>) for which it is relevant.

# 5. Results

## 5.1. Regexes

31 The data contain numerous instances of errors in regex scope or syntax, which result in matches not corresponding to the intended target (9).

(9)

a. (?<=s)c(?=[ie]) = /ʃ/ (eg *scena* > */sʃena/, cf. target /'ʃena/)

b. c(a) = /ka/

Verbose sets of rules (10a), inaccurate use of operators (10b) and sequential "find-and-replace" operations (10c) are also symptomatic of a struggle with regex syntax, although the transcription outcome is not necessarily compromised.

(10)

a.

sch = /sk/

sco = /sko/

scu = /sku/

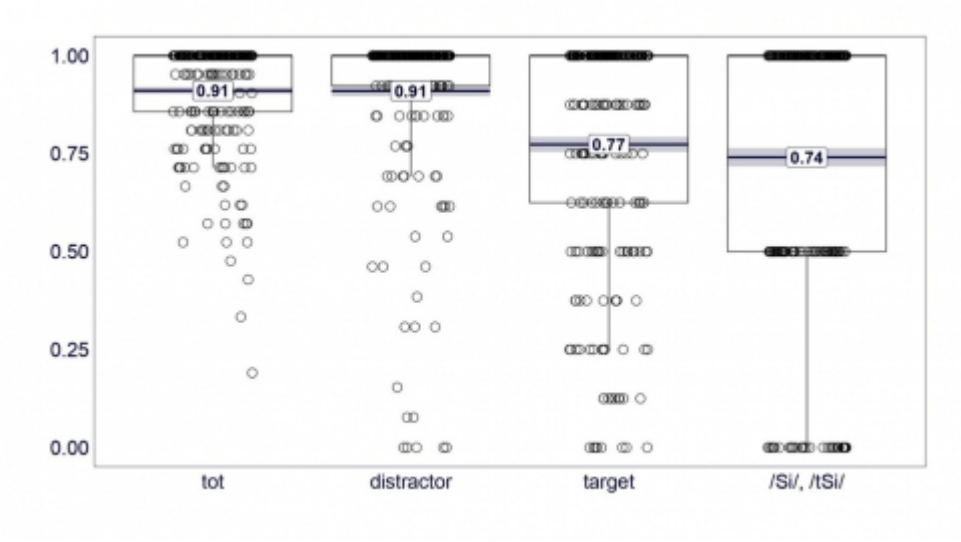
b.

(s)c(?=e)|(s)c(i)(?=a|o|u|e)|(s)c(?=i)(?!a|e|o|u) = /ʃ/

c.
sc(?=i|e) = /ʃ/
ʃi(?=a|o|u) = /ʃ/
c(?=i|e) = /ʧ/
ʧi(?=a|o|u) = /ʧ/

32 Participants' "find-and-replace" sequences proved on average quite effective, as shown in Graph 1. In addition to the standard elements of a boxplot, here and in the following graphs shaded areas represent credibility intervals, while the bold segment indicates the aggregated mean score, also reported in figures. Data points represent each participant's mean score. As expected, distractor items proved the least difficult. Target items (words comprising /ʃa/, /ʃo/, /ʃu/, /ʧa/, /ʧo/, /ʧu/) and <i> items produced somewhat lower scores. Between the latter two groups, the distribution of <i> item scores is more skewed towards lower values, which suggests that participants' regexes identified the target orthographic strings more often when <i> is followed by a vowel (and thus functions as a diacritic) than when it is followed by a consonant (and thus functions as both a diacritic and a vowel sound, cf. 4.2).

**Graph 1–Regex accuracy by item type.**




## 5.2. Transcription

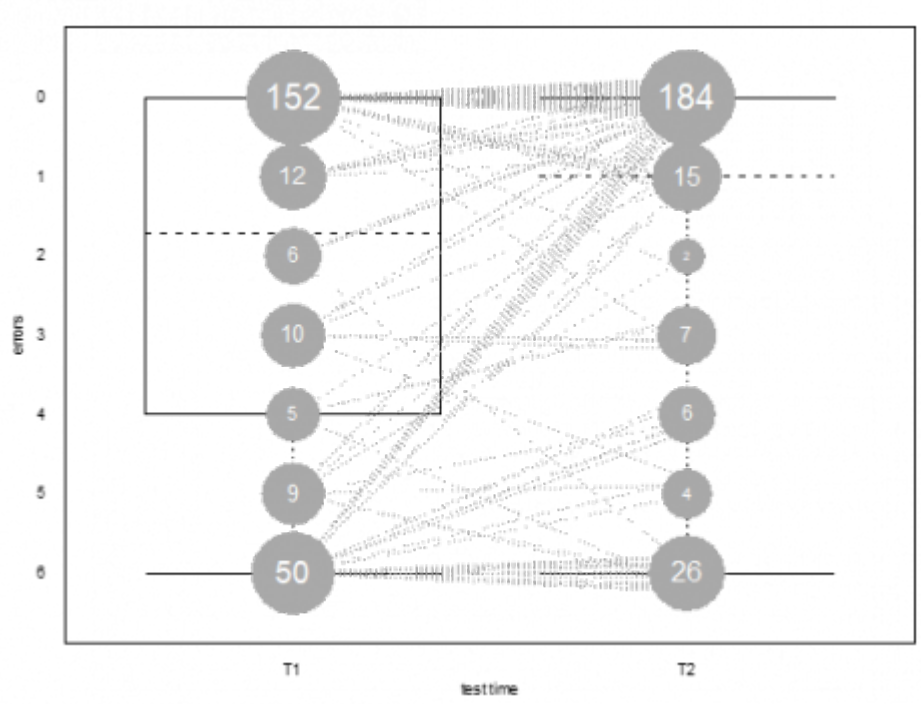
33 A macroscopic difference separates distractors and <i> items, both exhibiting a ceiling effect and negligible variation, from target items, processed with much lower accuracy at both test times (Table 6). For this reason, distractors and <i> items will not be considered any further in this section. Regarding target items, some improvement can be observed following the regex formalisation task.

**Table 6–Phonetic transcription scores by test time and item type.**

| Test time | Score type | mean | sd |
|---|---|---|---|
| T1 | Distractors | 1 | 0.02 |
| | <i> items | 1 | 0.06 |
| | Target items | 0.79 | 0.4 |
| T2 | Distractors | 1 | 0 |
| | <i> items | 1 | 0 |
| | Target items | 0.88 | 0.32 |

34 Graph 2 depicts the evolution of individual participants' scores between T1 and T2. The dashed segments identify the mean; the data point size is proportional to the number of participants (also indicated in figures inside the data point) who made the corresponding number of errors. In this respect, recall that at each test times, participants were presented with six target items, which–unlike distractors and <i> items–could potentially create transcription difficulties due to the diacritic role of the letter <i>. Dotted segments indicate participants' change in mean score (or lack thereof) between T1 and T2. As can be seen, scores at T2 are generally equal or higher than at T1. More specifically, the number of learners who scored 0% decreases from 50 at T1 to 26 at T2, while the number of participants scoring 100% increases from 156 to 189. It seems particularly noteworthy that 19 improved their score from 0% to 100%. Altogether, out of 244 participants, 59 improved their score, 16 scored worse (though often in a negligible manner) and 169 exhibited no change (of these, however, 142 scored 100% at T1).

**Graph 2–Number of errors in phonetic transcription by test time.**




## 5.3. Metalinguistic description

35 Numerous students provided accurate accounts of the orthographic rules in question (11a). It is quite striking that explicit mention is often made of the grapheme "<sc> followed by <a>, <o>, <u>, <h>" (11b), although this case is fully comprised in the more general rule "<c> followed by <a>, <o>, <u>, <h>." Other descriptions, though mostly correct, are sometimes inaccurate: in the most common formulation (11c), the graphic context is not sufficiently detailed, so that the two rules (I and ii) partially overlap:

(11)

a. the grapheme <ci> followed by <a, e, o, u> corresponds to /tS/; the grapheme <ci> not followed by a vowel corresponds to /tSi/;

b. the grapheme <sc> followed by <a>, <o>, <u> corresponds to /sk/;

c. /ʧ/ corresponds to <c> when it is followed by the vowels <e>, <i>;

/ʧ/ corresponds to <ci> when it is followed by the vowels <a>, <o>, <u>.

36 As expected, most erroneous answers ignore whether or not <i> is followed by a vowel or not (12a), or fail to identify a general rule (12b, 12c):

(12)

a. the sound /ʧ/ corresponds to the letter <c> followed by the vowels <i>, <e>;

b. the graphemes <sce>, <sci>, <scia>, <scio>, <sciu> correspond to /S/;

c. the sound /ʃ/ corresponds to the grapheme <sci>, except in words like "*uscita*". Some words that include the grapheme <ci> maintain <i> in their phonetic transcription. The experimental task may be thought to have exerted a positive effect when the learner's account at T2 successfully corrects or integrates its T1 counterpart (13).

(13)

T1

<c> followed by <e>, <i> corresponds to /ʧ/;

T2

<c> followed by <e>, <i> in turn not followed by vowels corresponds to /ʧ/;

<ci> followed by vowels corresponds to /ʧ/.

Finally, some participants (14) explicitly mention the regex that originated the metalinguistic reflection.

(14)

(?<!s)((ci(?=[aeou]))|c((?=e)|(?=(i(?![aeou])))))) = /ʧ/. <c> corresponds to /ʧ/ when it is not preceded by <s> and followed by <e, i> in turn not followed by <a, e, o, u>.

37 Only 43 students out of 255 provided accurate descriptions at T1. Out of the 212 who did not, however, 53 improved their analysis following the experimental task (Table 7).

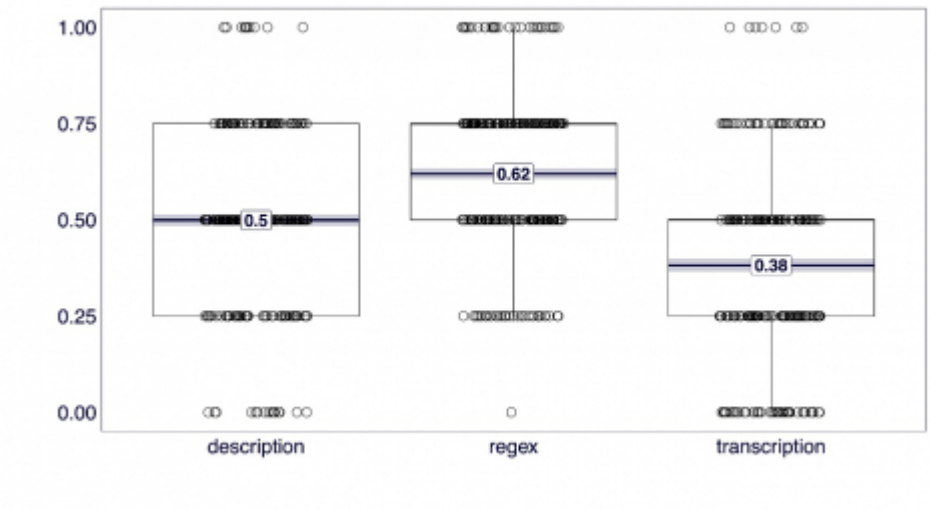
**Table 7–Accuracy of metalinguistic descriptions by test time.**

| | | T2 score | | total |
|---|---|---|---|---|
| | | 0 | 1 | |
| T1 score | 0 | 159 | 53 | 212 |
| | 1 | - | 43 | 43 |
| total | | 159 | 96 | 255 |

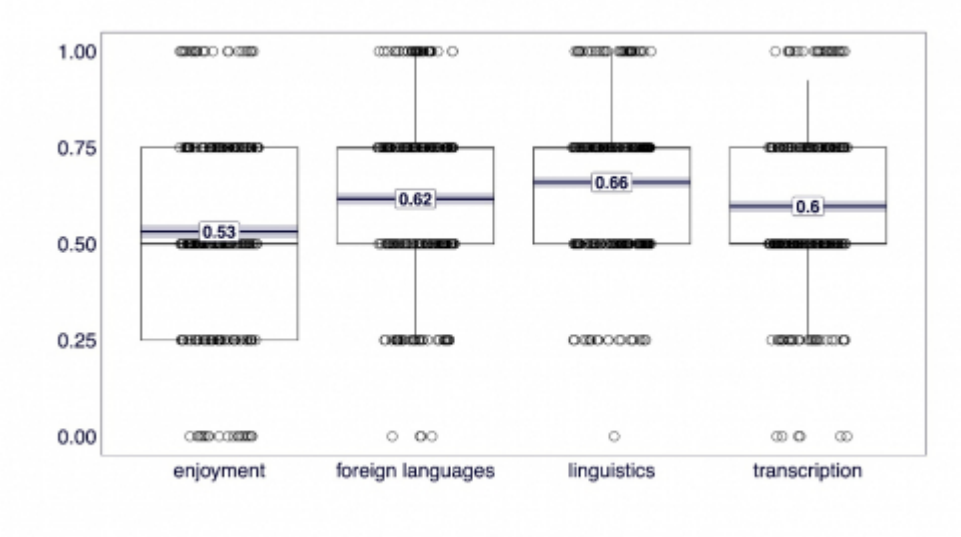
## 5.4. The participants' opinions

38 The participants' evaluations of the experimental task are presented in Graph 3 (perceived difficulty of the three components of the task) and Graph 4 (perceived usefulness of regexes for a variety of application fields).

**Graph 3–Difficulty of the three components of the task (0 = easy, 1 = hard).**




**Graph 4–Usefulness of regexes (0 = not useful, 1 = very useful).**




## 5.5. Inferential statistics

39 The participants' transcriptions were statistically analysed using a Bayesian multilevel model with binomial error structure and logit link function. Predictors included test time (T1 vs T2), type of target item (/ʃa/, /ʃo/, /ʃu/, /ʧa/, /ʧo/, /ʧu/), overall accuracy of the regexes produced by each participant (continuous, ranging from 0 to 1) and accuracy of metalinguistic description at each test time (binary). To account for individual variability, the model also included random intercepts for individual participants. Stabilising priors reflected a lack of commitment as to the direction and magnitude of the predictor effects.

40 Pairwise contrasts indicate a significant effect of test time (on the logit scale, like all following figures: mean = 2.14, sd = 0.21, 0.95 CI = 1–74 - 2.58), while no contrast concerning target type reached the significance threshold. The effect of overall regex accuracy proved positive and significant at the 95% threshold (mean = 2.7, sd = 0.68, 95% CI = 1.33–4.04), just like metalinguistic awareness (mean = 1.3, sd = 0.43, 95% CI = 0.49–2.15).

41 Concerning the participants' evaluation of the task, the perceived difficulty of phonetic transcription exhibits a negative, significant correlation with the participant's average transcription score (Pearson's product-moment correlation, t = -6.2, df = 242, p < 0.01, cor = -0.37, 95% CI = -0.47-0.26). No noteworthy correlation was found between the perceived difficulty of regex formalisation and regex average correctness. Finally, target item transcription scores weakly but significantly correlate with the accuracy of metalinguistic descriptions at T1 (t = 2.5, df = 236, p-value = 0.01, cor = 0.16, 95% CI = 0.04-0.28), but not at T2. This may be due to the fact that transcription accuracy only improved by 12%, while metalinguistic competence did so by 24%, which suggests that despite their increased awareness of the orthographic rules in question, some students still failed to adequately transcribe target items.

# 6. Discussion

42 The main results of the experiment may be summarised as follows. To start, there was an obvious difference in scores between distractor items and target items in which the letter <i> functions as a diacritic. The number of errors affecting the latter category decreased from 29% at T1 to 16% at T2. Most importantly, the students' descriptions of Italian orthography also improved over time. One may thus conclude that the reflection required to formalise an appropriate set of regexes had a positive impact on the participants' transcription accuracy and metalinguistic awareness. However, regexes are not part of the typical toolbox of foreign language students and take time to master.

Saturno (2020) asked a group of L1 Polish beginner students of L2 Italian to perform the same task than the one described in this paper, ie formalise a set of regexes describing the pronunciation of letter groups comprising a specific element of Italian orthography (the letter <c>, in that case). The participants were given one hour to familiarise with a few completely unknown regex operators. Leaving aside the fact that they were not native speakers of the target language (cf. infra), no positive effect of the task was observed because they evidently struggled with the regex syntax in the first place. Most of their attention was deviated from the target structure to the technical details of the new tool. Sufficient time and numerous exercises seem necessary for participants to sufficiently familiarise themselves with regexes to apply them to linguistic tasks.

43 Indeed, the course described in the present paper exclusively devoted about ten hours to this tool. It also included a set of exercises of increasing difficulty, of which the last one replicated exactly the structure of the experimental task, with the only exception of the target structure considered. In light of these considerations, it may seem legitimate to wonder whether the result is worth the pain of learning regexes. To answer this question, one might consider the manifold applications of regexes in a linguist's craft– and, partially, in that of a student of foreign languages, too. When working on languages with a transparent orthography, they can be very useful to automatically transcribe or transliterate large bodies of text with a few find-and-replace operations, thus eliminating the need to manually perform this time-consuming operation and the related possibility of human errors. The advantage becomes even more evident when regexes are used in association with a programming language to loop multiple find-and-replace operations. Further, regexes may be used to clean large bodies of data, originating for instance from an online survey. Finally, they have a wide range of data analysis applications (eg string manipulation within a spreadsheet or database) which, albeit not strictly linguistic themselves, may nonetheless prove useful to enrich a student's skill set. In this respect, too, the usefulness of regex skills is maximised by the knowledge of a programming language. However, this goal is rarely included in foreign language programmes and requires a substantial investment in terms of time and resources.

44 A further point concerns the utility of regexes for the study of foreign languages. Since the target structure belonged to the participants' native language, the present study considered exclusively the development of metalinguistic skills, with no language learning goal. However, the same approach could also be used as a heuristic method to inductively discover and memorise unfamiliar orthographic regularities. By elaborating and testing a set of regexes, participants are given the opportunity to test hypotheses and intuitions on the rules governing the phenomenon of interest. That was the research question that inspired the aforementioned experiment by Saturno (2020), which crucially took place at the very beginning of L2 Italian classes, when participants were hardly familiar with the target structure. On a more sophisticated level, the attempt to elaborate efficient, comprehensive regexes may again turn into an exercise in metalinguistic competence, potentially leading to the identification of variables related to linguistic features. In the present experiment, for instance, the most easily identifiable rule is that when <c> is followed by <a>, <o>, <u> (and <h>), it is pronounced /k/. This purely orthographic description can be interpreted linguistically by asking what these three vowels have in common. Linguistic knowledge should suggest that it is the "-front" trait, which in turns may prompt participants to (correctly) see the Italian sound /ʧ/ as the result of the palatalisation of /k/ when followed by a front vowel or semi-vowel. While the feasibility of this approach to develop metalinguistic knowledge in one's native language seems demonstrated by the present experiment, whether it also applies to foreign language learning is a question that remains open for further research.

45 The participants' opinions also provide useful information. It is noteworthy that the median for the perceived utility of the task for the purposes of linguistics and foreign languages is 4 out of 5 (0.75 in Graph 3), with a mean comprised between 3 (0.50) and

4 (0.75). The perceived difficulty of formalising regexes was evaluated in a very similar manner (Graph 3). One could conclude that the elaboration of regexes is perceived as useful for the consolidation of one's linguistic and metalinguistic skills, but also as quite demanding and potentially frustrating. Moreover, while phonetic transcription and the metalinguistic description of pronunciation rules are part of the skill set required of a student of foreign languages, regexes are not–in fact, only a negligible proportion of the participants were already (partly) familiar with this tool.

46 Concerning the generalisability of the results of this study, finally, the present task is obviously applicable only to languages characterised by a largely systematic sound-grapheme correspondence, like Italian or Polish. However, one should keep in mind the goal that inspired the introduction of regexes within a course addressing students of foreign language, which was not to teach the mechanism of Italian orthography within a language teaching course, but rather to stimulate the participants' metalinguistic reflection. Further, the participants were provided with the phonological transcription of the target words from the very beginning of the task, so that no additional knowledge of the target language was required to carry out the task in addition to the information provided. Together, these considerations suggest that the present task–or its adaptation to other orthographically transparent languages–can be used with foreign language students with any L1 background, while the constraints regarding the target language concern the transparency of its orthography, as mentioned. The training of the participants in view of the present experiment, for instance, included similar tasks based on Polish, a language they had no experience of.

# 7. Conclusion

47 In the experiment described in the present paper, a group of L1 Italian students of foreign languages were asked to formalise a set of regular expressions to automatically transcribe the Italian graphemes comprising the letter <i>, which, depending on the orthographic context in which it appears, may correspond to a vowel sound, or simply function as a diacritic. The results show that the task has a positive effect on the participants' transcription skills and metalinguistic awareness. One can thus conclude that the study of regexes, whose primary applications is ICT, can be effectively applied to the development of metalinguistic reflection skills, which in turn positively affect typically linguistic skills such as phonetic transcription. Even though the research question of the experiment can be answered affirmatively, one should consider that regexes are not typically part of the skillset of a student of foreign languages, and that developing sufficient proficiency in their technical details requires time and commitment. One could thus wonder whether the methodology described in this paper, albeit successful, can ultimately find a place in the training of foreign language students. Certainly, the potential of regexes to develop metalinguistic skills should be exploited in programmes with heavy emphasis on ICT, such as computational linguistics and digital humanities. Regarding less specialised recipients, regexes could probably find a place in ICT courses, especially in light of the general lack of specialised ICT skills among university students.

## *Bibliographie*


Andorno, C., Valentini, A., & Grassi, R. (2017). *Verso una nuova lingua : Capire l'acquisizione di L2* (1a edizione). UTET università.

Atom–A hackable text editor for the 21st Century (v1.46.0). (2020) https://atom.io/

Berruto, G. (2012). Che cosa vuol dire 'sapere una lingua'? Dai fonemi alle espressioni idiomatiche. In G. Bernini, C. Lavinio, A. Valentini, & M. Voghera (Eds.), Competenza e formazione linguistiche. In memoria di Monica Berretta (p. 27-53). Guerra. https://www.academia.edu/33262385/berruto_AItLA_BG.pdf.pdf

Berruto, G., & Cerruti, M. (2017). *La linguistica : Un corso introduttivo*. UTET università.

Calabrò, L. (2015). Il workshop di fonetica in italiano L2/LS. *Italiano LinguaDue*, *7*(1), 40–49. https://riviste.unimi.it/index.php/promoitals/article/view/5011
DOI : https://doi.org/10.13130/2037-3597/5011

Calaresu, E., & Dal Negro, S. (Eds.). (2018). *Attorno al soggetto. Percorsi di riflessione tra prassi didattiche, libri di testo e teoria*. AItLA. https://iris.unimore.it/handle/11380/1157434

Chang, W. (2014). extrafont : Tools for using fonts (0.17). http://CRAN.R-project.org/package=extrafont

De Masi, S., & Maggio, M. (2006). Gli studenti universitari : Quanta e quale grammatica conoscono. In D. Russo (Ed.), *Questioni linguistiche e formazione degli insegnanti* (p. 156-177). Franco Angeli.

Dragulescu, A., & Cole, A. (2020). xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files (0.65). https://CRAN.R-project.org/package=xlsx

Goyvaerts, J., & Levithan, S. (2009). *Regular expressions cookbook*. O'Reilly.

Guerrero, J. M. (2016). Sviluppare la competenza fonologica per acquisire la competenza ortografica. *Bollettino Itals*, *14*(66), 202-231. https://www.itals.it/sites/default/files/pdf-bollettino/novembre2016/bollettino_itals_66_guerrero.pdf

Katz, L., & Frost, R. (1992). The Reading Process is Different for Different Orthographies: The Orthographic Depth Hypothesis. In R. Frost & L. Katz (Eds.), *Advances in Psychology* (p. 67-84). Elsevier. https://www.sciencedirect.com/science/article/abs/pii/S0166411508627892?via%3Dihub
DOI : https://doi.org/10.1016/S0166-4115(08)62789-2
DOI : 10.1016/S0166-4115(08)62789-2

Landau, E., Lončarić, M., Horga, D., & Škarić, I. (1999). *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. Cambridge University Press.

Lavinio, C. (2011). (In)competenze metalinguistiche di base in laureati in lettere. *Italiano LinguaDue*, *3*(2). https://riviste.unimi.it/index.php/promoitals/article/view/1925
DOI : https://doi.org/10.13130/2037-3597/1925

McElreath, R. (2020). *Rethinking* [R]. Statistical Rethinking book package (R package version 2.13.). https://github.com/rmcelreath/rethinking

R Core team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing. http://www.R-project.org/

Saturno, J. (2020). Attività metalinguistiche su supporto informatico per lo sviluppo della competenza ortografica di apprendenti l'italiano L2. *Italiano LinguaDue*, *12*(2), 110-122. http://hdl.handle.net/10446/173175

Stan Development Team. (2019). Stan User's Guide. https://mc-stan.org/users/documentation/

Stan Development Team. (2020). RStan: The R interface to Stan (R package version 2.19.3.). https://cran.r-project.org/web/packages/rstan/vignettes/rstan.html

Torresan, P. (2010). Didattica per la costruzione della competenza fonetico-fonologica in lingua straniera. *Entre Lenguas*, *15*, 59-73. http://repositorioslatinoamericanos.uchile.cl/handle/2250/5209381

Viale, M. (2011). Le competenze grammaticali di base per l'università : Dati da un'esperienza didattica. In L. Corrà & W. Paschetto (Eds.), *Grammatica a scuola* (p. 137-152). Franco Angeli.

Wells, J. C. (1997). SAMPA computer readable phonetic alphabet. In D. Gibbon, R. K. Moore, & R. Winski (Eds.), *Handbook of Standards and Resources for Spoken Language Systems* (part IV, section B). Walter de Gruyter.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686
DOI : 10.21105/joss.01686

---

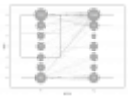
## *Notes*

1 IPA (international phonetic alphabet).

2 Originally "Informatica e dati linguistici."

3 In order to limit the complexity of the task, it was decided not to include words comprising the group <scie>, also pronounced /ʃe/, eg *scienza* /'ʃentsa/ "science."

## *Table des illustrations*

| | | |
|---|---|---|
|  | **Titre** | Figure 1–Using regexes to find abstract patterns in the text editor Atom. |
| | **URL** | http://journals.openedition.org/alsic/docannexe/image/6554/img-1.jpg |
| | **Fichier** | image/jpeg, 26k |
|  | **Titre** | Figure 2–Elaborating regexes through find-and-replace in the text file in Atom. |
| | **URL** | http://journals.openedition.org/alsic/docannexe/image/6554/img-2.jpg |
| | **Fichier** | image/jpeg, 30k |
|  | **Titre** | Graph 1–Regex accuracy by item type. |
| | **URL** | http://journals.openedition.org/alsic/docannexe/image/6554/img-3.jpg |
| | **Fichier** | image/jpeg, 115k |
|  | **Titre** | Graph 2–Number of errors in phonetic transcription by test time. |
| | **URL** | http://journals.openedition.org/alsic/docannexe/image/6554/img-4.png |
| | **Fichier** | image/png, 10k |
|  | **Titre** | Graph 3–Difficulty of the three components of the task (0 = easy, 1 = hard). |
| | **URL** | http://journals.openedition.org/alsic/docannexe/image/6554/img-5.jpg |
| | **Fichier** | image/jpeg, 99k |
|  | **Titre** | Graph 4–Usefulness of regexes (0 = not useful, 1 = very useful). |
| | **URL** | http://journals.openedition.org/alsic/docannexe/image/6554/img-6.jpg |
| | **Fichier** | image/jpeg, 114k |

## *Pour citer cet article*

*Référence électronique*
Jacopo Saturno, « ICT and metalinguistic competence: Developing phonetic transcription skills through regular expressions », *Alsic* [En ligne], Vol. 26, n° 3 | 2023, mis en ligne le 17 juin 2023, consulté le 26 juillet 2024. URL : http://journals.openedition.org/alsic/6554 ; DOI : https://doi.org/10.4000/alsic.6554

## *Auteur*

**Jacopo Saturno**
Jacopo Saturno is a researcher at the University of Bergamo, where he teaches courses on IT methods for the analysis of language data as well as General and Applied Linguistics. His research interests focus on L2 acquisition, with particular regard to Slavic nominal morphology and the role of input. He has also worked extensively on the possibility of intercomprehension within the Slavic and the Romance language families: in this respect, most of his research concerns the linguistic determinants of the phenomenon and its language teaching potential.
**Affiliation:** Università degli Studi di Bergamo.
**Email:** jacopo.saturno@unibg.it
**Web:** https://didattica-rubrica.unibg.it/ugov/person/3181
**Address:** Università degli Studi di Bergamo, Via Donizetti 3, 24129 Bergamo, Italy.

## *Droits d'auteur*