DOI: 10.17469/O2108AISV000016

DUCCIO PICCARDI, FABIO ARDOLINO

Gaming variables in linguistic research. Italian scale validation and a *Minecraft* pilot study

This paper deals with the concept of gamified science and its recent applications to the linguistic field. We argue that, albeit promising, this paradigm still lacks analytical tools to model the effects of the peculiar experimental setting on the results obtained. After a theoretical introduction to the User Engagement and Gaming Literacy constructs, we present two validated Italian translations of scales representing them. Lastly, we test these two gaming variables in a pilot study on the postvocalic realizations of /k t/ in the Florentine variety. Results show that both variables positively condition the production of non-continuants (i.e., emphasized words) but through different underlying mechanisms.

Keywords: gamification, User Engagement, Gaming Literacy, *Minecraft*, Florentine *gorgia*

1. *Gamification and science*

An interactive museum panel, an app for competing with other fitness-enthusiasts, an online game to explore a tourist destination: these are only a few examples of *gamified* activities nowadays familiar also to those who are less used to videogames. As pointed out in the analysis by Deterding, Dixon, Khaled & Nacke (2011), the notion of *gamification* – enjoying growing success in the last two decades – defines the adoption of typical videogame elements (design patterns, interface mechanics, rewards) in non-game contexts. The *ratio* under the gamification of traditionally non-game activities lies in the ease with which gamified experiences trigger the user's motivation, involvement, and entertainment, and, consequently, in the greater resolution with which the participants approach the completion of gamified tasks (Sailer, Hense, Mandl & Klevers, 2013).

In view of these opportunities, the development of gamified protocols in human sciences constitutes a natural outcome. Especially in data gathering, providing participants with entertaining tasks represents a crucial element to ensure the qualitative value of the collected data (Landers, Auer, Collmus & Armstrong, 2018). Gamified protocols have been applied in various research fields, generally bringing experimental advantages (Hamari, Koivisto & Sarsa, 2014).

As observed by Landers et al. (2018), however, the enthusiastic implementation of gamified approaches in many research scenarios has often underestimated the importance of critical reflection around the issues deriving from a naïve gamification of the investigation strategies (Rapp, Hopfgartner, Hamari, Linehan & Cena, 2019).

On the basis of these assumptions, this work aims at exploring how two of the main variables linked to the participant's performance in gamified protocols – namely, *User Engagement* (UE; § 2.1) and *Gaming Literacy* (GL; § 2.2) – interact with the collection of segmental data in a linguistic investigation focusing on a regional feature (the so-called *gorgia*, § 3.2) in the productions of Italian speakers from Florence. This experiment is described in § 3.4 and discussed in § 4. This work also provides a validation study of two Italian-translated tools for the quantification of UE and GL (§ 3.1). Before approaching the experimental discussion, the disciplinary field with which this study is collated (the gamification of linguistic research, § 1.1) will be addressed as well as the videogame used in the experiment: *Minecraft* (§ 1.2).

1.1 Gamified Linguistics

Linguistics has approached gamification relatively later in comparison to other humanistic fields. Nevertheless, the last decade has seen an increasing implementation of game elements in language studies and recently, the possibility of gathering linguistic data using gamified approaches has come to represent a trending topic explored in specialized publications and conferences (e.g., Hilton, Leemann, 2021). Applied Linguistics was one of the first fields to focus on the advantages of gamified approaches in improving language learning processes (Figueroa-Flores, 2015), followed by Cognitive Linguistics and Communication Studies. Already in the 2010s, the research by Yildirim, Narayanan & Potamianos (2011) constituted one of the first attempts to elicit conversational emotions in children using a gamified human-computer interaction. Thereafter, the study by Bos, Nissim (2015) comprised instead a first implementation of game-oriented tasks to facilitate the syntactic exploration of large digital corpora; the conclusion of the authors is that gamified approaches undoubtedly represent a noteworthy opportunity to simplify large-scale annotations.

The research developed by Leemann, Kolly, Purves, Britain & Glaser (2016) was one of the wider attempts to systematize the collection of speech data using gamified and participative methods. In Leemann, Schmid, Studer-Joho & Kolly (2018), a smartphone app was adopted to gather data regarding the regional variation of /r/ in Swiss dialects. The recent study by Duran, Lewandowski (2020) explicitly addressed the advantages of a videogame setting in an auditory discrimination task, specific measures having been adopted to quantify the subject's familiarity with videogames. Participants completed both a classic and a gamified experiment. A regression model aimed at explaining reaction times highlighted an interaction between test type and levels of previous gaming experience. Surprisingly, compared to the results of non-gamers, the reaction times of experienced players proved to be longer in the gamified test. This pattern suggests their sensitivity to deviations from common gaming conventions.

The prosodic study by Buxó-Lugo, Toscano & Watson (2018; see also Toscano, Buxó-Lugo & Watson, 2015) represents one of the main methodological references for the investigations carried out in this work. Buxó-Lugo et al. (2018) specifically

explored the predictor value exerted by participants' engagement – underpinned by a gamified approach – on the discriminability of prosodic cues. Aiming at experimentally manipulating the participants' engagement in communicative tasks, the authors adopted both a classic speech elicitation protocol (picture description) and a gamified one, consisting in the collaborative solution of simple puzzles in a *Minecraft* game environment. Higher levels of prosodic discriminability were observed in the gamified condition, and this was tentatively explained in relation to the greater spontaneity and emotional involvement triggered by gamification.

1.2 *Minecraft*: a playground for digital fieldwork

Minecraft (Persson, Bergensten, 2009) is a popular sandbox game periodically updated with new features. Similar to other sandbox games, *Minecraft* does not put the players in any predetermined ruleset, leaving them free to interact with the environment. The novelty introduced by this title is the *block*: in *Minecraft*, each scenario is composed of interactive cubic units. More than one hundred kinds of block can be found throughout the map, each one having distinct characteristics. At the cost of graphic realism, therefore, *Minecraft* allows a vast set of play possibilities.

The richness of the *Minecraft* experience has extended the fortune of this software far beyond the entertainment field. In 2016, a spin-off of the game was specifically developed for educational purposes (*Minecraft: Education Edition*; Mojang Studios, 2016; henceforth MEE). Compared to the original game, MEE introduces specific features to improve in-game teacher-student interaction. In line with its didactic goals, MEE is available free of charge for educational institutions and scholars, making this software an accessible tool for the design of gamified research protocols.

2. *Two variables for gamified linguistic research*

In § 1.1, we encountered two pivotal phenomena for the interpretation of gamified linguistic data. Gamification leads to participant engagement which, in turn, is expected to engender spontaneous speech (Toscano et al., 2015); moreover, it taps into the participants' knowledge of the digital medium, as gaming proficiency affects the responses to gamified inquiries (Duran, Lewandowski, 2020).

In Buxó-Lugo et al. (2018), engagement is conceptualized as a dichotomous label with a low (non-gamified task) and a high (gamified MEE task) level. However, no effort was made to detect engagement effects at the intra-task level: the exact linear relationships between discriminability of prosodic cues and participant engagement with the experiment remains therefore unexplored.

Duran, Lewandowski (2020) attempted to quantify individual proficiency by creating three pertinent variables: a binary *isGamer* (Yes/No) variable, the gaming frequency in days per week, and a score quantifying the number of game types played by the participants. Being presumably correlated, these three solutions struggle to

manifest autonomous interpretability. Indeed, there is no explanation as to why only the gaming score is retained through the proposed model reduction procedure.

Overall, it seems that these implementations may still be open to operational refinements. In the following sections, we try to find solutions to the construction of meaningful speaker-dependent variables for gamified linguistic experiments. Therefore, with the aim of selecting tools for their quantification, a theoretical contextualization is provided concerning the related constructs of *User Engagement* and *Gaming Literacy*.

2.1 User Engagement

The creation of engaging technologies is a major field of interest in human-computer interaction. While discussions date back to the second half of the 1980s and the 1990s (e.g., Laurel, 1991), UE has stolen the spotlight since the advent of the participative web, pervasive technologies, and an increase in user choices (Attfield, Kazai, Lalmas & Piwowarski, 2011). O'Brien, Toms (2008: 941) provided a working definition of UE as a "category of user experience characterized by attributes of challenge, positive affect, endurability, aesthetic and sensory appeal, attention, feedback, variety/novelty, interactivity, and perceived user control". In O'Brien (2016a), the author summarized her view on the major aspects of the construct, namely: *a)* UE is not one-dimensional, as it concerns affective, behavioral, and cognitive aspects; *b)* from the point of view of its temporality, UE manifests itself both during individual interactions and in subsequent states which condition the intention to re-engage; *c)* being bound to specific interactive events, UE takes contextuality seriously; and *d)* UE is not a dichotomous feature, its continuity being affected by user-, system-, and interaction-specific characteristics.

This summary highlights the shortcomings of treating UE as categorical and entirely task-dependent in gamified linguistic protocols (Buxó-Lugo et al., 2018). Variationist concerns add urgency to the need for a thorough investigation of the UE effects on linguistic data. Studies suggest that specific socio-demographic groups may react differently to engaging settings: Yang, Du, Wang & Wu (2019), observed a UE interaction with gender, while Di Gangi, Wasko (2016) observed one with age. Overall, these data suggest that the behavioral consequences of UE are not uniform across the population, so that adding interactions with social variables may unveil important data patterns.

Over the years, several UE quantification strategies have been devised. Among these, the UE Scale (UES) has received considerable attention, counting more than forty implementations in the most disparate contexts just a few years after its first release (O'Brien, 2016b). The UES was developed in O'Brien, Toms (2010) as a 31-item tool, divided into *perceived usability*, *aesthetics*, *focused attention*, *felt involvement*, *novelty*, and *endurability*. Eventually, two main issues related to the scale became evident (O'Brien, 2016b; O'Brien, Cairns & Hall, 2018): firstly, researchers perceived the UES as too cumbersome; secondly, reexaminations of the full scale suggested a rearrangement of its six factors to four (Wiebe, Lamb, Hardy

& Sharek, 2014). O'Brien et al. (2018) addressed these two points by validating a new four-factor structure, the components of which were re-labeled as *aesthetic appeal* (AE), *focused attention* (FA), *perceived usability* (PU), and *reward* (RW). At the same time, the authors proposed a reduction to three items per factor, resulting in a 12-item short form.

To sum up, even though UE scales developed for the specific context of videogaming do exist (Abbasi, Ting & Hlavacs, 2017), the handiness of the UES short form and its many applications in heterogeneous research environments may favor its implementation in gamified linguistic research.

2.2 Gaming Literacy

In a broad sense, GL is part of the digital literacies and, more generally, the so-called *new literacies*. This label is a postmodernist attempt at challenging the monolithic and hierarchical acceptation of literacy as the mastery of reading and writing (Lankshear, Knobel, 2011). The internet plays a pivotal role in this decentralization process (Reinhardt, Warner & Lange, 2014: 161): thus, since their first definition (Gilster, 1997), digital literacies have been one of the most successful instantiations of this construct, being about "mastering ideas, not keystrokes" (Gilster, 1997: 15; Bawden, 2008: 18 and ff.). (Video)gaming literacy inherited the holistic nature of its hypernyms. This is a clear point in some of its most renowned definitions: for example, Gee (2003: 45-46) interlinked GL with a model of critical learning involving the decoding of a semiotic space and the transfer of knowledge to other domains and Zimmerman (2009) equated it to acquiring a set of cognitive, creative, and social skills in order to engage with the world and interpret its meanings. Nonetheless, mere system skills are still considered as core components of GL (Rosenberg, 2011: 20), and articulated models of literacy usually represent them in dedicated subcategories (e.g., Kringiel, 2012: 637).

Duran, Lewandowski (2020) focused on this aspect of literacy. Indeed, controlling the relative easiness of the participant-setup interaction is essential in any evaluation of research results (see, e.g., Rama, Black, Van Es & Warschauer, 2012). Mutatis mutandis, this point reflects classic concerns in sociolinguistics. Commenting on the manipulation of style attempted through the Labovian interview, Baugh (2001) observed that the stylistic variation triggered through reading sessions heavily depends on the participant's literacy (in the traditional acceptation). If the hurdles experienced by the participant in interacting with the reading setup become too tough, the experiment ultimately fails. In addition to that, gaming-proficient participants can rely on their rooted habits and strategies, altering the intended effects of the manipulation. In another gamified perception experiment (a forced-choice categorization test), Lewandowski, Duran (2019) observed that the expected positive effect of mental flexibility, assessed through a Simon test, on correct categorizations was observable in the responses by female participants only. While gaming proficiency per se was not quantified, it was

hypothesized that, being reputedly more gaming proficient, male participants may not rely on the benefits of higher mental flexibility.

Given the complex nature of GL, the development of scales going beyond the core component of skill is still far behind the refinement of UE. To the best of our knowledge, Rosenberg's (2011) dissertation is the only systematic attempt at developing and validating a GL scale. This study led to an extensive 47-item, 5-factor scale, which is not suited to variable construction in non-GL-focused research (Rosenberg, 2011: 87). Interestingly enough, the first 22-item factor (*Information and Systems Management*), which included the core components of GL concerning mental and technical proficiency, drastically outperformed the others in terms of explained variance. Unfortunately, further revisions and reductions to the scale were never provided. This study also presented some adaptations to the gaming contexts of other reliable analytic tools. For example, devising a correlational counterpart to his GL scale for the sake of nomological validation, Rosenberg (2011) adopted the skill, challenge, and play scales designed in Novak, Hoffman (1997) to investigate the manifestation of flow in web users. Flow was conceptualized by Csikszentmihalyi (1975: 35 and ff.) as the achievement of the optimal experience in performing an activity. Rosenberg (2011: 27) argued that GL and flow share a conceptual linkage, namely that literate players are expected to reach high flow levels. In Rosenberg's validation analysis, this assumption was generally met. In order to pinpoint a compact tool representing a construct that is at least closely tied to the core aspects of GL, we inspected the results concerning the main *Information and Systems Management* factor of Rosenberg's original scale. This factor was most significantly correlated with and the best predictor of the five-item skill scale by Novak, Hoffman (1997). Thus, the strength of this theoretical relation may be considered reliable and it can be provisionally suggested that this manageable scale can be of some use in quantifying GL in gamified linguistic research[1].

3. *Gamified phonetics in Italian: scale validation and a Florentine pilot study*

This paragraph expounds the experimental component of this work. Firstly (§ 3.1), a validation study is presented of the Italian translations of the UES short form (O'Brien et al., 2018) and the skill scale adapted from Novak, Hoffman (1997) by Rosenberg (2011), interpreted here as a rough approximation to GL. Then, we test for the effectiveness of these scales by proposing an expansion to the *Minecraft*

[1] An objection often raised concerning the formulation of items assessing GL is their genericity about videogame types. It is useful to report the argument of Klecka, Johnston, Bowman & Green (2021: 5) about conceptualizing GL through the total time spent videogaming. Klecka and colleagues observe that – in consideration of the vast articulation of today's videogame market – cumulative time measurements can't represent the knowledge of a subject on specific genres. While Novak, Hoffman's (1997) skill scale does not face this problem, future GL research should aim to find a good compromise between genre representativity and handiness of the scales.

linguistic agenda. While previous studies have focused on prosody (Buxó-Lugo et al., 2018), we try to export this method to the analysis of the segmental behavior of participants/gamers. Previous MEE studies hypothesized that engagement led participants to spontaneousness, which in turn makes them *a)* want to communicate effectively and *b)* produce emotion-driven utterances. Even though emotions were interpreted as not decisive, it should be noted that both phenomena converge on enhanced prosodic discriminability. From the segmental viewpoint, the effects of engagement may not be so straightforward for both phonetic and sociolinguistic reasons. While the facilitation of comprehension implies hyperarticulation (cf. Lindblom, 1990), spontaneity promotes enhanced variation (Barry, Andreeva, 2001), which also includes extreme reduction (e.g., Adda-Decker, Lamel, 2018). Labovian theorization of spontaneity (Labov, 2006: 64-65) promotes such complexification. In variationist terms, spontaneity is a condition that reduces the attention paid to speech and promotes more informal productions. This acceptation does not refer to the hyper/hypo continuum, as the "spontaneous" form is shaped by the indexical values attributed to specific variants[2].

In § 3.2, we will introduce the object of our MEE experiment (§ 3.4), the Florentine *gorgia* (lit. 'throat'; lenition of postvocalic plosives). This widespread phenomenon is characterized by reduced articulatory effort (Marotta, 2008), and the exceptions to its occurrence still need to be properly explained. The design of the gamified experiment will be described in a separate section (§ 3.3).

3.1 Italian UE and GL scales. Validation and internal structure

Since our study involves Italian participants, the experimental scales adopted to quantify UE (O'Brien et al., 2018) and GL (Rosenberg, 2011) were opportunely translated and re-validated. Preliminarily to the study, a three-step translation process was put in place to ensure the closeness of the Italian scales to the original ones. After the first translation into Italian, a bilingual computer engineer with experience in usability testing was asked to re-translate the items into English: items judged as not sufficiently similar to the original ones were reoperated; the process was repeated until agreement was met. The final items were randomized and arranged in a questionnaire, which was filled in by our participants using 5-point Likert scales (see the *Appendix*) after the completion of the MEE task (§ 3.3).

[2] This cursory reference to speech spontaneity oversimplifies a major terminological and conceptual issue in linguistics and should be intended as merely expedient to underline the complexity of analyzing a communicative context in which engagement and spontaneity do not unequivocally point to a single direction in the hypo/hyper continuum. See Mereu, Vietti (2021) for a recent take on spontaneity. Also, in this section we do not make reference to any sociolinguistic acceptations of style other than Labovian attention paid to speech. Of course, other models of style, such as the audience design (i.e., style shifting is conditioned by the features of the audience) or speaker-centered approaches (i.e., style as agentive identity construction in a given context; Hernández-Campoy, 2016) might be of interest in the field of gamified phonetics. In fact, we have noted some traces of individual agency modulating our experimental results(§ 5, *d*).

The trustworthiness of the Italian scales has been tested for reliability with widely adopted approaches. To improve the representativeness of the scales and obtain more accurate validation coefficients, the tests were carried out on the values returned by a pool of Italian participants (N = 32) which is larger than the Florentine subset analyzed in § 3.4 (N = 21). The results of the validation tests are summarized as follows.

The UES translated in this study consists of 12 items, grouped according to four distinct factors: *aesthetic appeal* (AE), *focused attention* (FA), *perceived usability* (PU), and *reward* (RW), each one comprising three items. A *Cronbach's alpha* test was adopted to verify the internal reliability for each of the factors, returning acceptable values in three out of four cases: FA, $\alpha = .65$; PU, $\alpha = .77$; RW, $\alpha = .7$. Concerning the AE factor, the returned alpha value ($\alpha = .58$) almost reaches the lower acceptability threshold of 0.6: in view of the relatively small number of items, this value appears to be equally suitable for the study (Cortina, 1993). The global alpha returned by the scale ($\alpha = .8$) is acceptable. Once the validity of the translated scale had been confirmed, a *Confirmatory Factor Analysis* (CFA) was carried out to verify the correspondence between the internal structures of the original and the translated scales (*lavaan* R package: R Core Team, 2020; Rosseel, 2012). The fit of the model based on the original UES four-factor structure returns the following indices: $\chi^2(48) = 75.74$, $p = .007$; *Comparative Fit index* (CFI) = .77; *Tucker Lewis index* (TLI) = .68; *Root Mean Square Error of Approximation* (RMSEA) = .134, $p = .02$. None of the computed fit indices achieve robust acceptability; however, due to the absence of abnormal values compared to the traditional thresholds and the potential effect induced on the coefficients by small samples (e.g., Chen, Curran, Bollen, Kirby & Paxton, 2008) we decided to accept the model for the aims of this study, deferring a more thorough inspection of the scale structure until the sample is sufficiently large.

The same validation analyses were carried out on the GL scale. Since no internal factor structure is envisaged for this tool (all the items refer to a one single factor), the test was modulated accordingly. The *Cronbach alpha* test carried out on the translated scale returns a robust index of $\alpha = .85$, confirming the internal reliability of the Italian item battery. Following Novak, Hoffman (1997), a *Principal Component Analysis* was performed to explore the composition of the model: differently from the original skill scale, in which a single component exceeds the variance proportion minimum threshold of 10%, the translated scale returns two significant components, although with markedly unbalanced variance proportions (PC1 = 63%, PC2 = 16%). Since only the aggregate measure of the scale is under analysis in this study, a similar result does not seem to impact the trustworthiness of the obtained scale.

3.2 A gamified stress test for the *gorgia*. Brief history and working hypothesis

From the phonetic viewpoint, the *gorgia* is probably the most discussed linguistic feature of the Florentine (and Tuscan) way of speaking, counting an extensive

academic tradition focused on its debated origins (Izzo, 1972). The *gorgia* is a series of consonant lenition phenomena in postvocalic position, including the variable weakening of the phonological voiceless plosives (/k t p/; e.g., ['seθa], *seta*, 'silk'). The feature has a peculiar areal distribution stemming from the city of Florence, in which the phenomenon is most systematic; overall, the velar voiceless plosive is the preferred target, followed by /t/ and /p/, considering both parameters of frequency of occurrence and lenition degree (Giannelli, Savoia, 1979-1980).

Since the first acoustical analyses commented in Giannelli (1976) and Giannelli, Savoia (1978), researchers have struggled to provide categorical descriptions of the extreme variability of the *gorgia* phonetic outputs, which are not limited to fricatives (Marotta, 2001). By investigating a corpus of 938 potential *gorgia* realizations uttered by six young Florentine speakers, Sorianello (2001) realized that *a)* the most common Florentine *gorgia* realizations of /k t p/ are [ɦ θ ɸ] and *b)* non-continuant productions are overall rare (13.7%): among this small pool of occurrences, only 13 tokens (lenis and voiced plosives) pertain to /k t/. This pervasiveness of continuants differs from the distributions found in other Tuscan towns (Pisa, Siena), in which peculiar facets of non-continuity have been reported as "noticeably closed" fricatives (Giannelli, Savoia, 1978), preaspirated fricatives (Stevens, Hajek, 2005), semifricatives and homorganic affricates (Marotta, 2001). These variants consist of a more or less silent closure phase followed by a long release friction without a concentrated burst; from the perceptual viewpoint, they can be confused with plosives *stricto sensu*.

In order to gain a better understanding of the apparently rare Florentine non-continuants, previous research tried to pinpoint the circumstances of their occurrence. Giannelli (1976) and Giannelli, Savoia (1978: 43) reported very small percentages (3-5%) of Florentine non-continuants, used in fast and precise styles to stress specific words. Moreover, the authors referred to contextually motivated productions of Italian-like plosives, which may end up being hypercorrected through hyperarticulation (i.e., aspirated voiceless plosives). Other studies have stressed the roles of hyperarticulation and social factors in the selection of non-continuant variants in Florence. In another analysis of 500 potential *gorgia* contexts uttered by three Florentine speakers during a spontaneous dialogue, Sorianello (2003) did find a small percentage of stops. The author explains that "the occurrence of a voiceless stop is context-dependent: it occurs only in hyperarticulated or emphasized words" (Sorianello, 2003: 3082). Crucially, the data presented in Villafaña Dalcher (2008) goes way beyond this fragmentary evidence. Of her 660 *gorgia* contexts, 35% (more than double the number in Sorianello, 2001) was produced as non-continuants. These included stops (18%), fricated (or leaky) stops[3] (13%) and semifricatives (4%) and were distributed among the three points of articulation (/k/: 14% of 231; /t/: 53% of 232; /p/: 39% of 197). While socio-indexical factors (such as the enactment

[3] Fricated stops are here defined as produced with an incomplete closure (with visible noise) – burst – release temporal sequence. See Villafaña Dalcher (2006: 69) for a reference spectrogram (Florentine speech). In the corpus presented in Villafaña Dalcher (2008), they were mostly /t/ allophones.

of supralocality) played a significant role in this pattern (Villafaña Dalcher, 2008: 310), the author also remarks that her experimental setup (read speech and the mention of Florentine *Italian* as the primary interest of the researcher) may have conditioned the emergence of standard-like productions.

To sum up, while the Florentine variety manifests pervasive weakening of postvocalic voiceless plosives, researchers also noted that there is room (albeit small) for non-continuity. However, to date, these variants have always been considered as collateral in the search for weakening. Turning back to the MEE gamified protocol, two elements of particular interest emerged in the review of *gorgia* studies: non-continuants *a)* are produced in order to stress specific words through hyperarticulation and *b)* are task-sensitive or, in other words, they can presumably be elicited through experimental manipulation. The MEE protocol is designed to trigger spontaneous speech, aiming at effective communication between players (§ 3). In the context of *gorgia* productions, UE may represent an ambiguous predictor of speaker behavior: on the one hand, the achievement of (sociolinguistic) spontaneousness may lead to a higher number of informal local variants (weakened *gorgia* tokens); on the other, engagement may motivate speakers to do well in the communicative MEE game through an enhancement of the phonetic cues of potential disambiguation (hyperarticulated non-continuant tokens). Conversely, the expected effects of core GL are more straightforward: as it is in the interest of expert gamers to communicate effectively in order to succeed they may make greater articulatory effort (i.e., through non-continuants). Therefore, in the following section, we will test the hypothesis that core GL positively affects the number of non-continuant variants of phonological voiceless plosives produced by Florentine speakers during a MEE session. UE is expected to follow this trend, but with less incisiveness. In fact, the predicted enhancement of discriminability (Buxó-Lugo et al., 2018) may be dampened by the presence of (sociolinguistically) spontaneous *gorgia* tokens. Moreover, by testing the interaction between GL and UE, we hypothesize that UE positively conditions the number of non-continuant productions, but only for those speakers manifesting a high GL. This follows the principle that a sufficient level of literacy is required in order to be sensitive to the intended experimental manipulation (cf. Baugh, 2001). In other words, from the viewpoint of an expert player who is more competent in "reading" the rules of the game, engagement could be less ambiguously related to the deployment of effective communication strategies (a sort of "literate engagement").

3.3 A brief parenthesis on level design

The MEE level presented to participants consists of a small game arena shaped as a castle. At the start of the game, participants find themselves in front of the entrance and are free to familiarize themselves with the key controls. They receive all the game instructions from a non-playable character positioned near the main door. The goal of the level is to cross the castle from the first to the last of six consecutive communicating halls, which are connected by blocked doors, unlockable through

the activation of levers in the correct sequence. Each hall is characterized by a similar set of elements, with some differences from one hall to another. The MEE developer toolset was inspected to retrieve all the objects corresponding to words whose Italian names include postvocalic /k t/. Due to the limitations of the game resources, a small set of 5 candidates was deployed to manipulate the environmental differences, while other elements were inserted as distractors. In detail, these elements are: *a)* an enclosure containing a sheep (It. *pe̲co̲ra*), the color of the wool varying from hall to hall; *b)* four frames representing vegetables (a potato, It. *pa̲ta̲ta*; a carrot, It. *caro̲ta*; a watermelon, It. *co̲co̲mero*), the kinds and order of which vary across the halls, with at least one of the three stimuli being repeated every time; *c)* a set of four mural tiles, encoding a distinct color sequence for each hall (distractors); *d)* four pots, containing flowers the colors of which vary across the halls (distractors); and *e)* a carpet (It. *tappe̲to*) with a distinct pattern depending on the hall.

The only way to discover the correct sequence of activation of the levers for each hall is to describe the exact content of the room to the researcher. Once all the information has been received, the experimenter identifies the room on a solution sheet and communicates the way out of the puzzle: at this point, the player can pull the levers to unlock the exit door. When the participant crosses the last castle door, another NPC notifies the end of the experience.

3.4 The *gorgia* at play

3.4.1 Participants and setting

In 2021, 21 Florentine speakers (11 females, 10 males; age M(SD) = 40.09(14.9)), took part in the MEE experiment. In most cases, participants (friends and acquaintances) were recruited by the first author. The sessions took place in three rooms in two Municipalities near Florence (Scandicci, Greve in Chianti).

3.4.2 Materials

This gamified experiment used the MEE level described in § 3.3.

3.4.3 Procedure

In order to reduce potentially confounding factors of sociolinguistic nature, this stress test for the pervasiveness of the Florentine *gorgia* was performed by pairing our local informants with an in-group researcher (first author)[4]. Participants sat without headphones in front of a pc with the MEE level already running. The

[4] Two issues – non-experimental in nature – strongly impacted the protocol designed for this study, namely *a*) the impossibility of using two MEE accounts at the same time and in the same game arena due to the restriction of the MEE free use plan and *b*) the impossibility, due to COVID-19 restrictions, of gathering together a greater number of participants within the same space, in order to perform the experimental task without the active intervention of the experimenter. Future redefinitions of the protocol should have as a primary goal the reduction of the researcher presence in the experimental task, through *a*) the generation of a totally in-game interaction between experimental actors and *b*) an elicitation method based on the interaction between two informants (a player and a helper, bearing the solution sheet), who are free to interact to solve the puzzles.

researcher assisted the participants in navigating the very first part of the level, until they entered the castle. Then, the researcher stepped back diagonally and positioned himself at approximately two meters from the participant, facing the opposite direction while holding the solution sheet. While a complete fixation of the researcher's behavior during the sessions was unfeasible, he scrupulously stuck to a small set of communicative rules: *a)* never intentionally solicit word repetitions; *b)* never steer the participant's lexical choices toward the intended lexeme; *c)* avoid any unnecessary productions of the target words, while using paraphrases instead. All the tokens produced during a violation of these points were excluded from the analysis. The sessions were recorded through a Zoom H5 diagonally positioned in front of the participant (.wav 24 bit, 48 KHz). As the task instructions never mentioned language issues, participants were completely free to select any variant of their linguistic repertoire.

3.4.4 Analysis

The acoustic analysis of the recorded sessions was performed in *Praat* (Boersma, Weenink, 2021). Target words were manually segmented, and their respective durations were extracted to compute a local speech rate measure (log(syllables/seconds)). Then, the postvocalic contexts of interest were inspected and coded as follows: 1, continuant (or deleted) segment (presence of a "canonical" Florentine *gorgia*); 0, non-continuant segment (absence of a "canonical" Florentine *gorgia*). The latter category included aspirated plosives (i.e., with voice onset time/VOT above 30 ms long; Cho, Ladefoged, 1999: 223), short-lag VOT plosives, leaky plosives (see § 3.2), and homorganic affricates. A series of statistical analyses were run to test our working hypotheses. A *Generalized Linear Mixed-Effect* model (*lme4* R package: Bates, Maechler, Bolcher & Walker, 2015; the argument *family* was set to "binomial") was built trying to predict the binary *gorgia* response from speaker's UE (M(SD): 3.71(0.77)) and GL (M(SD): 2.97(1.38)) (main predictors); speaker's sex, age, sophistication, and regionality, as well as consonant type (/k/ *vs.* /t/), stress position (stressed *vs*. post-stress syllable) and speech rate (controls). Sophistication was organized into three groups (adapted from Hudson, 1996), which were grounded on both occupation and education level: sophisticated (white-collar graduates, N = 6), rough (blue-collar non-graduates, N = 9), and mixed (intermediate scenarios, N = 6). A participant's regionality (or degree of localness) was conceptualized through an adaptation of the Regionality Index (RI: Chambers, Heisler, 1999). Our RI consisted of four questions regarding the participant's own birthplace and that of their parents, their place of residence, and relocations. The researcher coded the answers as follows: In the Province of Florence = 0; In another Tuscan Province = 1; In another Italian Region = 2; In another country = 3. We averaged these scores in order to obtain individual RI values (M(SD): 0.33(0.34)). Participant and word were added as random factors. Then, we investigated all the interactions between the main predictors and the control

variables. Interaction selection was performed adapting the guidelines of Stuart-Smith, Sonderegger, Rathcke & Macdonald (2015), i.e., by keeping all the interactions which proved to be significant both when inspected in isolation and together with the other variables. Lastly, a model reduction procedure was manually performed by pruning non-significant variables until the best compromise between Akaike and Bayesian Information Criterion (AIC/BIC) values was met (Hay, Foulkes, 2016).

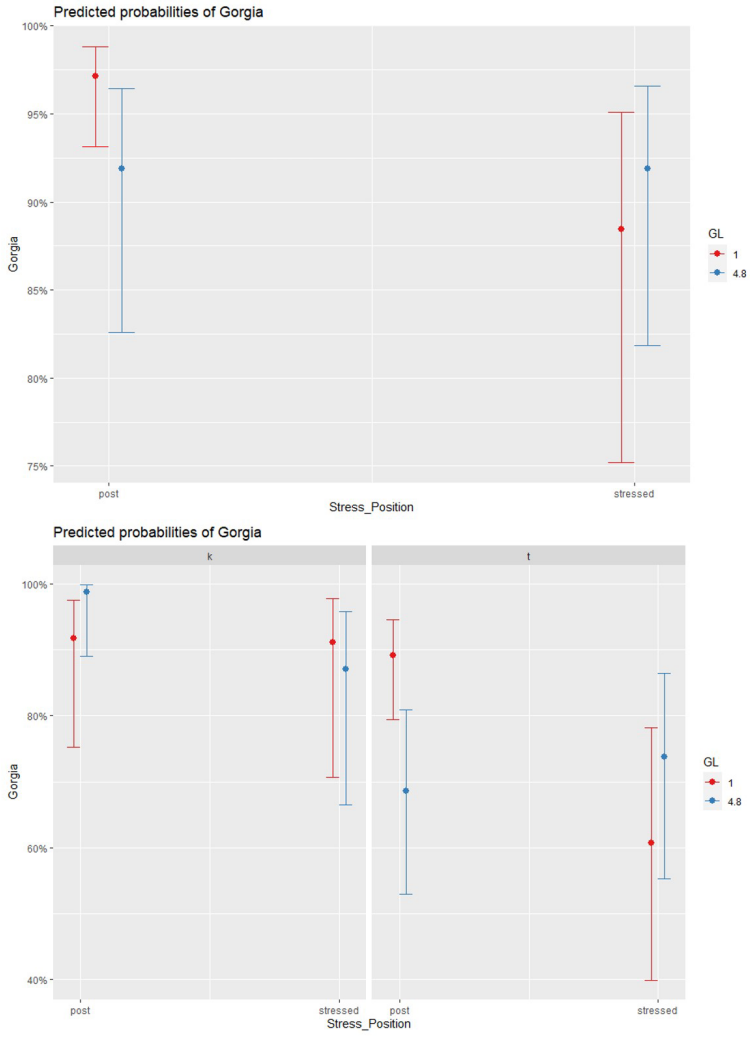
3.4.5 Results

799 contexts of analysis were produced by our participants. These included 165 *carota*, 75 *cocomero*, 155 *patata*, 155 *patata*, 115 *pecora* and 134 *tappeto*. The differences in token availability depended on the structure of the halls, lexical alternatives (e.g., *anguria vs*. *cocomero*) and the individual tendency to produce repetitions that did not violate the communicative rules expounded in the *Procedure* section (contexts of analysis per participant: M(SD) 38.05(6.04)). 182 tokens were labelled as non-continuants, 166 of which pertained to /t/ (27.26% of total /t/ productions), 16 to /k/ (8.42%); 115 to post-stress positions (20.21%) and 67 to stressed positions (29.13%); 65 to male participant speech (17.10%) and 117 to females (27.92%); 62 to "sophisticated" participant speech (25.31%), 43 to "mixed" participants (19.28%) and 77 to "rough" participants (23.26%).

Table 1 - *Summary of the final gorgia model calculating the changes in log odds of producing a continuant (or deleted) segment per unit change of the predictor. Recall that the other categories in the reported dichotomies are Consonant (/k/) and Stress Position (Post-stress). Observations: 799. Random factor: Participant (variance: .58; std. dev.: .76)*

	Estimate	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	4.62	1.39	3.32	<.001***
Consonant (/t/)	-1.51	0.31	-4.94	<.001***
GL	-0.29	0.15	-1.89	.058
Speech Rate	0.90	0.38	2.40	.016*
Stress Position (Stressed)	-1.88	0.48	-3.91	<.001***
*Stress Position (Stressed)*GL*	0.39	0.14	2.71	.007**
UE	-0.67	0.27	-2.47	.014*

Figure 1 - *Comparison of the interactions between GL and stress position (above), and GL, stress position and consonant (below). Evidently enough, the former is a generalization of the /t/ distribution highlighted in the latter*



The first full model had convergence issues, which we resolved by removing the "Word" factor with random intercept variance close to 0[5] and imposing a *bobyqa* optimizer with a high number of function evaluations. Our pipeline led us to a model with UE (negative effect), speech rate (positive effect), consonant (/t/ = negative effect), and an interaction between GL and stress position. The interaction between UE and GL was not significant (est. -.23; std. error: .2;

[5] As an anonymous reviewer pointed out, given the very small pool of stimuli, a certain level of collinearity between "Word" and "Consonant" is not surprising.

p = .26), even though it highlighted the expected pattern (i.e., UE conditions the production of non-continuants in the speech of participants with high GL, while low-GL participants are mostly unaffected). Since one of the few studies noting stress effects on the Florentine *gorgia* also investigated their interactions with plosive point of articulation (Sorianello, Bertinetto & Agonigi, 2005), we fitted a triple interaction between GL, plosive type, and stress position. This addition was significant (est. 1.14; std. error: .49; p = .019); however, while a slight improvement of AIC was noted (Δ>2), the triple interaction was unacceptable in terms of BIC (Δ>10). Fig. 1 compares the plot of the two interactions (*sjPlot* package: Lüdecke, 2021), highlighting that, while overall the participants with high levels of GL were unaffected by a positive effect of post-stress position on the *gorgia* continuant productions, this pattern ultimately pertains to the /t/ dataset only. Tab. 1 shows the final model.

4. *Discussion*

The two validated Italian scales for UE and GL were put into action through a pilot gamified inquiry on the emergence of non-continuant allophones of /k t/ in postvocalic position in the Florentine variety. Exceptions to continuant *gorgia* productions may happen to emphasize specific words (Sorianello, 2003). For this reason, we inferred that a game based on the effectiveness of player communication might well be the ideal context for the observation of these variants. Moreover, we hypothesized that skilled (and, secondarily, more engaged) players enact a clearer and more effective communication with more non-continuants, and that having previous experience with videogames "unlocks" an additional effect of being engaged with the experimental setting on the production of non-continuity. While these assumptions were only partially met, UE and GL did influence the *gorgia* patterns, hinting at principles of regularity coherent with our theoretical expectations.

Our percentages of non-continuant allophones can be considered an intriguing compromise between the read speech of Villafaña Dalcher (2008) and the (semi-and) spontaneous conversations of Sorianello (2001, 2003). While /k/ is indeed pervasively produced as continuant, /t/ manifests a noticeable number of non-continuant alternatives. The Florentine *gorgia* prefers velar segments: this is one of the few, if not the only (Bertinetto, Sorianello & Ricci, 2007), distributional feature to be constantly retrieved in the relevant literature, and our statistical analysis makes no exception. However, this pattern does not explain why, across different studies, the percentages of non-continuity seem to vary the most in the realizations of /t/[6]. Apparently, being the middle ground in the front/back axis,

[6] Recall that Sorianello (2003) noted about 3% of non-continuant /t/ (less than the non-continuant /k/ allophones), while Villafaña Dalcher (2008) presented a staggering 53% (far more than the non-continuant /p/ allophones!) and our study has a 27.26% of non-continuity.

which conditions the degree of weakening in postvocalic position, /t/ is an optimal locus for the enactment of socially or communicatively driven variation. In fact, arguments have been advanced pointing at phonological and articulatory factors pressuring Tuscan /k/ towards lenition and /p/ towards (relative) preservation (i.e., consonant strength scales: Sorianello, 2001: 82; asymmetries in the Italian phonological system, which does not present back fricatives; combination of high intraoral pressure and reduction in tongue body constriction in V/k/V sequences: Marotta, 2008: 250-251, Villafaña Dalcher, 2006), so that /t/ realizations may be relatively more free to vary[7].

Speech rate proved to be a significant predictor of the *gorgia*: at higher rates, the probability of observing a continuant production increases. This effect is coherent with the gradient decrease in segment duration, which was observed across the multiple degrees of weakening of the *gorgia* variants (Marotta, 2001; Sorianello, 2001[8]), but is at odds with the relation between fast speech styles and non-continuity which was mentioned in Giannelli (1976) and Giannelli, Savoia (1978). We can tentatively argue that these two viewpoints can be reconciled by adopting a multidimensional acceptation of linguistic style (e.g., Moore, 2004) as an agentive practice of self-expression, which is de facto individual and linguistically complex, thus eluding the statistical patterns. In our corpus, there were indeed instances of "fast and precise" non-continuants; however, these occurrences were apparently very rare and speaker-specific, so that they did not constitute a problem for the robustness of the expected effect directionality.

Lastly, previous studies on the *gorgia* reported mixed results concerning stress position. With respect to Pisan data, Marotta (2001: 37-38) hypothesized that non-continuity is favored in stressed syllables. However, her analysis of variance failed to find support to this idea. In two computational studies based on Florentine data, Bertinetto et al. (2007) did not manage to find any stress effects on the *gorgia*, while Sorianello et al. (2005) observed a small trend confirming Marotta's (2001) hypothesis, although non-continuity was favored in stressed syllables mostly in the case of dentals and bilabials. Our descriptive percentages seem to be in line with this phenomenon; however, keep in mind that our lexical pool was extremely limited because of MEE's technical boundaries. In our set of stimuli, stressed syllables were represented as part of repeated CV sequences (i.e., *patàta*, *cocòmero*). As patterns of consonant lenition are sensitive to repetition avoidance (cf. Walter, 2007), any acritical generalization of this distribution should be avoided. Be that as it may, in our statistical model, the "Stress position" variable was ultimately part of a significant interaction with "GL", which will be discussed next.

[7] An alternative view on the issue highlights the cross-linguistic properties of the "coronals", which justify their being the best target for variation in general (e.g., Barbera, Barth, Frassinelli & Lapesa, 2009: 114).

[8] See Cohen Priva, Gleason (2020) for clarifications on the causal relationships between speech rate, segment duration, and lenition.

This short presentation of the linguistic variables conditioning the *gorgia* was expedient to comment on the effects of the two gaming scales. Most importantly, both GL and UE proved to be relevant predictors of participant behavior during the gamified experiment. This bodes well for their usefulness in unveiling linguistic patterns and promotes their adoption in other contexts of inquiry. With respect to GL, our initial hypothesis was drastically complexified through interaction testing. Rather than a general trend towards hyperarticulation, it seems that GL motivates specific communicative strategies involving the stress position variable. Fig. 1 shows that, while less skilled participants manifested a prosodic pattern favoring non-continuant productions in stressed syllables and continuant *gorgia* in post-stress position, the choices of more skilled gamers were stable across prosodic contexts. At first glance, this resembles Lewandowski, Duran's (2019) failure to observe the expected effects of the experimental manipulation on the behavior of proficient subjects, who can presumably rely on alternative contextual resources. However, although statistically redundant, the triple interaction plot helps us to get a clearer picture of the specific behavioral difference between low- and high-GL participants. While both categories tend to have high percentages of continuant *gorgia* for postvocalic /k/, the predictions are more variegated for /t/. In the /t/ section of the plot, the major GL discrepancy lies in the production of tokens in post-stress position: high-GL speakers have a lower probability of producing a continuant *gorgia* in this context. Recall that post-stress /t/ was represented by tokens in word-final syllables (*tappeto*, *carota*, *patata*), while the /k/ counterpart was post-stress but word-internal (*pecora*). In Sorianello et al. (2005: 352), the authors conclude that weak *gorgia* allophones are most typical of phonological plosives in post-stress, prosodic-unit final syllables. A tentative explanation of the GL pattern could posit that the frequent struggles with the gaming interface caused hesitations in the speech of less proficient players, favoring prosodic-unit final productions and, consequently, word-final weakening. Conversely, the fluency of more skilled players did not hinder their overall communicative planning and eventual hyperarticulation. While an in-depth analysis of the prosodic structure of our sessions could clarify this issue, it should be stressed that, from the linguistic viewpoint, GL adapts to the generalities of the Florentine system and underlines local differences in the behavior of the participants.

The same cannot be stated for UE, which per se positively predicted a higher number of non-continuants in general. This result is quite surprising if we interpret the equation *engagement* : *spontaneity* in a sociolinguistic sense. A positive engagement in light-hearted, Florentine communication should have favored the local, informal *gorgia* continuant variants; conversely, the spontaneousness which is encapsulated by our UE seems to be related to the personal involvement in an effective communication. One way to break this impasse is to detach our variant selection process from the sociolinguistically meaningful continua of formality and localness (Villafaña Dalcher, 2008). While these axes are indeed pertinent to the *gorgia* variation, it should be kept in mind that the previous short descriptions

of Florentine allophonic non-continuity do include pragmatic features (i.e., word emphasizing) which exist beyond them. In other words, a sociolinguistically spontaneous, casual Florentine conversation may well include postvocalic non-continuants: as long as they are used for stressing specific words, these allophones do not evoke any indexical incoherence. Therefore, we may argue that, with respect to the *gorgia*, UE is free to trigger phonetic spontaneousness, i.e., a removal of "cruise control" (Barry, Andreeva, 2001: 64) in the selection of hyper and hypo forms. Overall, our tentative interpretations suggest that, while GL describes how participants interact with the gamified setup in a technical sense, UE recounts their personal involvement, resulting in more general linguistic behaviors: in fact, its statistical significance is autonomous and not tied to another predictor.

Lastly, the lack of significance of the hypothesized interaction between GL and UE can be explained in relation to the extremely accessible mechanics of MEE. After a few initial struggles, even our 70-year-old participants who had never played a videogame in their entire lives successfully understood the rules and completed our levels. Therefore, GL did not appreciably filter the content of participant UE. In any case, as the related plot did point towards the expected trend, we might expect to find significant interactions in more complex gamified experiments.

5. *Conclusions: current limitations and future directions*

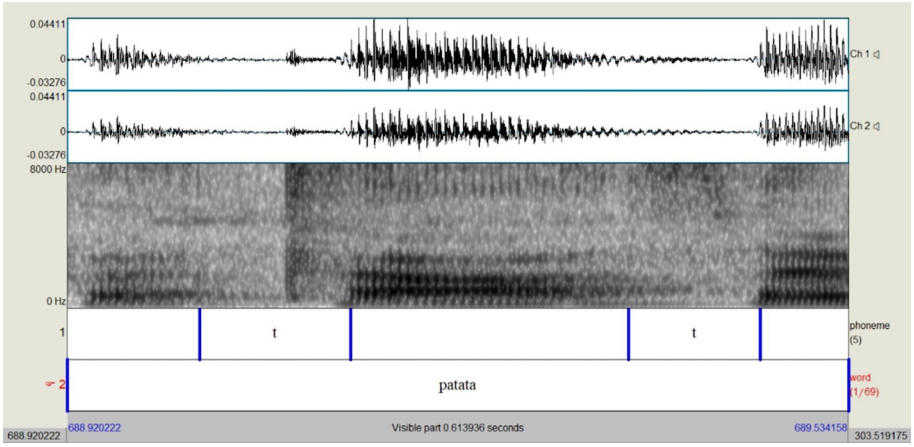
In this study, we proposed a validation of the Italian short-form UES (O'Brien et al., 2018) and gaming skill scale (Rosenberg, 2011), which is here considered a rough approximation to GL. Although reliability tests are already acceptable, as our pool of participants is very limited, future replications of our results using larger Italian datasets may be called for. A larger number of responses would benefit also the internal structure analyses, the results of which are still far from optimal. At present, we tentatively recommend adopting these scales in their aggregate forms only and invite Italian researchers to experiment with their factor structures in other contexts.

The MEE experiment was succeful in proving the usefulness of the two gaming variables in linguistic research. GL and UE represent two different types of conditioning on linguistic production: the former seems to be mostly related to local adjustments to technical difficuties, while the latter was linked to general behavioral patterns because of the level of personal involvement in the task. We argue that the extreme accessibility of MEE somehow dampened the expected role of GL: future research should experiment with more complex communicative games such as *Keep talking and nobody explodes* (Steel Crate Games, 2015). Moreover, we are aware that our implementation of GL is rudimental and still in need of an effective summarization of its multifaceted nature (also concerning the diversity of contemporary videogame genres), and we stress that gamified research could vastly benefit from a dedicated scale construction process.

The experiment has proved to be of dialectological interest as well, as it is the first to explicitly focus on the production of non-continuant variants of voiceless plosives in postvocalic position in the Florentine variety. The percentages of non-continuants, which were higher than in other spontaneous corpora (Sorianello, 2001, 2003), and the positive effect of UE on the preference for non-weakened allophones are small hints favoring the idea that these variants are produced for word emphasizing in specific discourse structures. However, this postulate should be refined in several ways: *a)* in order to reconcile the partially conflicting correlates of spontaneity and engagement (§§ 3, 3.2) with our UE results, we argued that non-continuity might well find its place in specific occurrences of spontaneous Florentine speech. Future research might be interested in quantifying spontaneity separately from engagement by counting the occurrences of specific indicators (e.g., laughs, disfluencies, etc.; Mereu, Vietti, 2021); *b)* the set of lexical targets was extremely limited and engendered ambiguity as the effect of multiple variables could not be effectively separated. While this was due to an intrinsic limitation of MEE, the research community should actively search for a workaround to this issue; *c)* a control, non-gamified experimental condition could help define the task-specific role of UE and GL by comparing the production of the same set of words in different circumstances (cf. Buxó-Lugo et al., 2018); *d)* while this pilot study has focused on a binary distinction between continuants and non-continuants, we suggest that drawing more precise coordinates of the emergence of the individual allophones could be an intriguing dialectological task. For example, to the best of our knowledge, this study was the first to experimentally replicate the Florentine postvocalic aspirated plosives observed in Giannelli, Savoia (1978). These allophones exclusively occurred in the speech of a retired teacher (see Fig. 2), who could have preferred extreme hyperarticulation in order to index precision (Drummond, Schleef, 2016). Are these allophonic selections idiosyncratic, or can we expect to find more systematic social patterns? How does indexicality merge with more general pragmatic roles of hyperarticulation? *e)* In-depth analyses of the prosodic and temporal aspects of the gamified sessions (hesitations, total completion time, etc.) could provide more complete pictures of the factors conditioning the produced speech from a segmental viewpoint.

Overall, gamified linguistic research is still at its title screen. Through the validation of gaming-specific variables, this field can enter new engaging levels of linguistic analysis.

Figure 2 - *Patata ('potato') uttered by a 71-years-old, female, white-collar participant with high UE (4.25) and very low GL (1). The first /t/ is produced as aspirated plosive (VOT: 51 ms, covering the 41.8% of the total segment duration)*



Bibliography

Abbasi, A.Z., Ting, D.H. & Hlavacs, H. (2017). Engagement in games: developing an instrument to measure consumer videogame engagement and its validation. In *International Journal of Computer Games Technology*, 10, 1-10. https://doi.org/10.1155/2017/7363925

Adda-Decker, M., Lamel, L. (2018). Discovering speech reductions across speaking styles and languages. In Cangemi, F., Clayards, M., Niebuhr, O., Schuppler, B. & Zellers, M. (Eds.), *Rethinking Reduction*. Berlin/Boston: De Gruyter Mouton, 101-128. https://doi.org/10.1515/9783110524178-004

Attfield, S., Kazai, G., Lalmas, M. & Piwowarski, B. (2011). Towards a science of user engagement (Position paper). In *WSDM '11: Proceedings of the fourth ACM international conference on Web search and data mining*, Hong Kong, China, 9-12 February 2011, 1-8.

Barbera, M., Barth, M., Frassinelli, D. & Lapesa, G. (2009). Là dove il [ʃi] suona. Un'analisi parametrica delle sibilanti nell'Italiano di Carrara. In Romito, L., Galatà, V. & Lio, R. (Eds.), *La fonetica sperimentale: metodo e applicazioni. Atti del IV Convegno Nazionale AISV*. Torriana: EDK, 105-127.

Barry, W., Andreeva, B. (2001). Cross-language similarities and differences in spontaneous speech patterns. In *Journal of the International Phonetic Association*, 31(1), 51-66. https://doi.org/10.1017/S0025100301001050

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. In *Journal of Statistical Software*, 67(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Baugh, J. (2001). A dissection of style-shifting. In Eckert, P., Rickford, J.R. (Eds.), *Style and sociolinguistic variation*. Cambridge: Cambridge University Press, 109-118. https://doi.org/10.1017/CBO9780511613258.007

Bawden, D. (2008). Origins and concepts of digital literacy. In Lankshear, C., Knobel, M. (Eds.), *Digital Literacies: Concepts, Policies and Practices*. New York: Peter Lang, 17-32.

Bertinetto, P.M., Sorianello, P. & Ricci, I. (2007). Sulla gerarchia dei fattori che governano la gorgia. Un'applicazione dell'algoritmo di decisione C&RT. In Giordani, V., Bruseghini, V. & Cosi, P. (Eds.), *Scienze Vocali e del linguaggio – Metodologie di valutazione e risorse linguistiche. Atti del III Convegno Nazionale AISV*. Torriana: EDK, 167-180.

Boersma, P., Weenink, D. (2021). Praat [Computer Program]. Phonetic Sciences, University of Amsterdam. Version 6.2, retrieved November, 15 2021 from https://www.fon.hum.uva.nl/praat/.

Bos, J., Nissim, M. (2015). Uncovering noun-noun compound relations by gamification. In Megyesi, B. (Ed.), *Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA 2015)*, Vilnius, Lithuania, 11-13 May 2015, 251-255.

Buxó-Lugo, A., Toscano, J.C. & Watson, S.G. (2018). Effects of participants' engagement on prosodic prominence. In *Discourse Processes*, 55(3), 305-323. https://doi.org/10.1080/0163853X.2016.1240742

Chambers, J.K., Heisler, T. (1999). Dialect Topography of Quebéc City English. In *Canadian Journal of Linguistics*, 44(1), 23-48. https://doi.org/10.1017/S0008413100020703

Chen, F., Curran, P.J., Bollen, K.A., Kirby, J. & Paxton, P. (2008). An Empirical Evaluation of the Use of Fixed Cutoff Points in RMSEA Test Statistic in Structural Equation Models. In *Sociological Methods & Research*, 36(4), 462-494. https://doi.org/10.1177/0049124108314720

Cho, T., Ladefoged, P. (1999). Variation and Universals in VOT: Evidence from 18 Languages. In *Journal of Phonetics*, 27, 207-229. https://doi.org/10.1006/jpho.1999.0094

Cohen Priva, U., Gleason, E. (2020). The causal structure of lenition: A case for the causal precedence of durational shortening. In *Language*, 96(2), 413-448. https://doi.org/10.1353/lan.2020.0025

Cortina, J.M. (1993). What is coefficient alpha? An examination of theory and applications. In *Journal of applied psychology*, 78(1), 98-104. https://doi.org/10.1037/0021-9010.78.1.98

Csikszentmihalyi, M. (1975). *Beyond Boredom and Anxiety. The Experience of Play in Work and Games*. San Francisco/London: Jossey-Bass.

Deterding, S., Dixon, D., Khaled, R. & Nacke, L. (2011). From Game Design Elements to Gamefulness: Defining Gamification. In *Proceedings of the 15th International Academic MindTrek: Envisioning future media environments*, Tampere, Finland, 28-30 September 2011, 9-15.

Duran, D., Lewandowski, N. (2020). Demonstration of a Serious Game for Spoken Language Experiments-GDX. In Lukin, S.M. (Ed.), *Proceedings of the LREC 2020 Workshop Games and Natural Language Processing*, Marseille, France, 11-16 May 2020, 68-78.

Di Gangi, P.M., Wasko, M.M. (2016). Social Media Engagement Theory: Exploring the Influence of User Engagement on Social Media Usage. In *Journal of Organizational and End User Computing*, 28(2), 53-73. http://doi.org/10.4018/JOEUC.2016040104

Drummond, R., Schleef, E. (2016). Identity in Variationist Sociolinguistics. In Preece, S. (Ed.), *The Routledge Handbook of Language and Identity*. London: Routledge, 50-65.

Figueroa-Flores, J. (2015). Using Gamification to Enhance Second Language Learning. In *Digital Education Review*, 27, 32-54.

Gee, J.P. (2003). *What videogames have to teach us about learning and literacy*. New York/Basingstoke: Palgrave Macmillan.

Giannelli, L. (1976). *Toscana*. Pisa: Pacini.

Giannelli, L., Savoia, L.M. (1978). L'indebolimento consonantico in Toscana. I. In *Rivista Italiana di Dialettologia*, II, 23-58.

Giannelli, L., Savoia, L.M. (1979-1980). L'indebolimento consonantico in Toscana. II. *Rivista Italiana di Dialettologia*, III-IV, 38-101.

Gilster, P. (1997). *Digital Literacy*. New York: John Wiley & Sons.

Hamari, J., Koivisto, J. & Sarsa, H. (2014). Does Gamification Work? A Literature Review of Empirical Studies on Gamification. In *Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS '14),* Waikoloa, HI, USA, 6-9 January 2014, 3025-3034.

Hay, J., Foulkes, P. (2016). The evolution of medial /t/ over real and remembered time. In *Language*, 92(2), 298-330. https://doi.org/10.1353/lan.2016.0036

Hernández-Campoy, J.M. (2016). *Sociolinguistic styles*. Chichester: Wiley-Blackwell.

Hilton, N.H., Leemann, A. (2021). Editorial: using smartphones to collect linguistic data. In *Linguistics Vanguard*, 7(1), 1-7. https://doi.org/10.1515/lingvan-2020-0132

Hudson, R.A. (1996). *Sociolinguistics*. 2nd edition. New York: Cambridge.

Izzo, H.J. (1972). *Tuscan & Etruscan*. Toronto: University of Toronto Press.

Kringiel, D. (2012). Learning to play: Video game literacy in the classroom. In Fromme, J., Unger, A. (Eds.), *Computer games and new media cultures: A handbook of digital games studies*. Berlin: Springer, 633-646. https://doi.org/10.1007/978-94-007-2777-9_40

Klecka, H., Johnston, I., Bowman, N.D. & Green, C.S. (2021). Researchers' commercial video game knowledge associated with differences in beliefs about the impact of gaming on human behavior. In *Entertainment Computing*, 38, 1-11. https://doi.org/10.1016/j.entcom.2021.100406

Labov, W. (2006). *The Social Stratification of English in New York City*. 2nd Edition. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511618208

Landers, R.N., Auer, E.M., Collmus, A.B. & Armstrong, M.B. (2018). Gamification science, its history and future: definitions and a research agenda. In *Simulation & Gaming*, 49(3), 315-337. https://doi.org/10.1177/1046878118774385

Lankshear, C., Knobel, M. (2011). *New Literacies. Everyday Practices and Social Learning*. 3rd Edition. Maidenhead/New York: Open University Press.

Laurel, B. (1991). *Computers as Theatre*. Reading: Addison-Wesley.

Leemann, A., Kolly, M.-J., Purves, R., Britain, D. & Glaser, E. (2016). Crowdsourcing Language Change with Smartphone Applications. In *PLoS ONE*, 11(1), 1-25. https://doi.org/10.1371/journal.pone.0143060

Leemann, A., Schmid, S., Studer-Joho, D. & Kolly, M.-J. (2018). Regional Variation of /r/ in Swiss German Dialects. In *Proceedings of Interspeech 2018*, Hyderabad, India, 2-6 September 2018, 2738-2742.

Lewandowski, N., Duran, D. (2019). Individual differences in implicit attention to phonetic detail in speech perception. In *Proceedings of Interspeech 2019*, Graz, Austria, 15-19 September 2019, 2255-2259.

Lindblom, B. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In Hardcastle, W.J., Marchal, A. (Eds.), *Speech Production and Speech Modelling*. Dordrecht: Springer, 403-439. https://doi.org/10.1007/978-94-009-2037-8_16

Lüdecke, D. (2021). sjPlot: Data Visualization for Statistics in Social Science. [R package] Version 2.8.8, retrieved June, 11 2021 from https://CRAN.R-project.org/package=sjPlot.

Marotta, G. (2001). Non solo spiranti. La gorgia toscana nel parlato di Pisa. In *L'Italia Dialettale*, 62, 27-60.

Marotta, G. (2008). Lenition in Tuscan Italian (Gorgia Toscana). In Brandão de Carvalho, J., Scheer, T. & Ségéral, Ph. (Eds.), *Lenition and fortition*. Berlin: Mouton de Gruyter, 235-271. https://doi.org/10.1515/9783110211443.2.235

Mereu, D., Vietti, A. (2021). Dialogic ItAlian: the creation of a corpus of Italian spontaneous speech. In *Speech Communication*, 130, 1-14. https://doi.org/10.1016/j.specom.2021.03.002

Mojang Studios (2016). Minecraft: Education Edition (PC Version) [Videogame]. Microsoft. Version 1.17.31, retrieved December, 15 2020 from https://education.minecraft.net/.

Moore, E. (2004). Sociolinguistic Style: A Multidimensional Resource for Shared Identity Creation. In *The Canadian Journal of Linguistics*, 49(3/4), 375-396. https://doi.org/10.1017/S0008413100003558

Novak, T.P., Hoffman, D.L. (1997). Measuring the flow experience among web users. Paper presented at the Interval Research Corporation.

O' Brien, H.L., Cairns, P. & Hall, M. (2018). A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. In *International Journal of Human-Computer Studies*, 112, 28-39. https://doi.org/10.1016/j.ijhcs.2018.01.004

O'Brien, H.L. (2016a). Theoretical perspectives on user engagement. In O'Brien, H.L., Cairns, P. (Eds.), *Why Engagement Matters*. Cham: Springer, 1-26. https://doi.org/10.1007/978-3-319-27446-1_1

O'Brien, H.L. (2016b). Translating theory into methodological practice. In O'Brien, H.L., Cairns, P. (Eds.), *Why Engagement Matters*. Cham: Springer, 27-52. https://doi.org/10.1007/978-3-319-27446-1_2

O'Brien, H.L., Toms, E.G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. In *Journal of the American Society for Information Science and Technology*, 59(6), 938-955. https://doi.org/10.1002/asi.20801

O'Brien, H.L., Toms, E.G. (2010). The development and evaluation of a survey to measure user engagement. In *Journal of the American Society for Information Science and Technology*, 61(1), 50-69. https://doi.org/10.1002/asi.21229

Persson, M., Bergensten, J. (2009). Minecraft (PC Version) [Videogame]. Mojang. Version 1.16, retrieved March, 03 2020 from https://www.minecraft.net/.

R Core Team (2020). R: A language and environment for statistical computing. [Computer program] Version 4.0.5., retrieved March, 05 2021 from https://www.r-project.org/.

Rama, P., Black, R., Van Es, E. & Warschauer, M. (2012). Affordances for second language learning in World of Warcraft. In *ReCALL*, 24(3), 322-338. https://doi.org/10.1017/S0958344012000171

Rapp, A., Hopfgartner, F., Hamari, J. & Cena, F. (2019). Strengthening gamification studies: Current trends and future opportunities of gamification research. In *International Journal of Human-Computer Studies*, 127, 1-6. https://doi.org/10.1016/j.ijhcs.2018.11.007

Reinhardt, J., Warner, Ch. & Lange, K. (2014). Digital games as practices and texts: New literacies and genres in an L2 German classroom. In Pettes Guikema, J., Williams, L. (Eds.), *Digital Literacies in Foreign and Second Language Education*. San Marcos: CALICO, 159-177.

Rosenberg, K.A. (2011). Gaming Literacy: Construct Validation and Scale Construction. Master of Arts thesis, University of South Florida.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. In *Journal of Statistical Software*, 48(2), 1-36. https://doi.org/10.18637/jss.v048.i02

Sailer, M., Hense, J., Mandl, H. & Klevers, M. (2013). Psychological perspectives on motivation through gamification. In *Interaction Design and Architecture(s) Journal*, 19, 28-37.

Sorianello P., Bertinetto P.M. & M. Agonigi (2005). Alle sorgenti della variabilità della 'gorgia' fiorentina: un approccio analogico. In Cosi, P. (Ed.), *La misura dei parametri. Aspetti tecnologici ed implicazioni nei modelli linguistici. Atti del I Convegno Nazionale AISV.* Torriana: EDK, 327-362.

Sorianello, P. (2001). Un'analisi acustica della 'gorgia' fiorentina. In *L'Italia Dialettale*, 62, 61-94.

Sorianello, P. (2003). Spectral characteristics of voiceless fricative consonants in Florentine Italian. In Solé, M.J., Recasens, D. & Romero, J. (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain, 3-9 August 2003, 3081-3084.

Stevens, M., Hajek, J. (2005). Spirantization of/p t k/ in sienese Italian and so-called semi-fricatives. In *Proceedings of Eurospeech/Interspeech 2005*, Lisbon, Portugal, 4-8 September 2005, 2893-2896.

Steel Crate Games (2015). Keep talking and nobody explodes (PC version). Steel Crate Games. Version 1.0, retrieved January, 18 2022 from https://keeptalkinggame.com/.

Stuart-Smith, J., Sonderegger, M., Rathcke, T. & Macdonald, R. (2015). The Private Life of Stops: VOT in a Real-time Corpus of Spontaneous Glaswegian. In *Laboratory Phonology*, 6, 505-549. https://doi.org/10.1515/lp-2015-0015

Toscano, J.C., Buxó-Lugo, A. & Watson, D.G. (2015). Using game-based approaches to increase level of engagement in research and education. In Dikkers, S. (Ed.), *Teachercraft*. Pittsburg: ETC Press, 139-151.

Villafaña Dalcher, C. (2006). Consonant weakening in Florentine Italian: An acoustic study of gradient and variable sound change. Ph.D. Dissertation, Georgetown University.

Villafaña Dalcher, C. (2008). Consonant weakening in Florentine Italian: a cross-disciplinary approach to gradient and variable sound change. In *Language Variation and Change*, 20(2), 275-316. https://doi.org/10.1017/S0954394508000021

Walter, M.A. (2007). Repetition in Human Language. Ph.D. Dissertation, MIT.

Wiebe, E.N., Lamb, A., Hardy, M. & Sharek, D. (2014). Measuring engagement in video game-based environments: investigation of the user engagement scale. In *Computers in Human Behavior*, 32, 123-132. https://doi.org/10.1016/j.chb.2013.12.001

Yang, H., Du, H.S., Wang, L. & Wu, T. (2019). The influence of social support networks on health conditions via user engagement: gender as moderator. In *Journal of Electronic Commerce Research*, 20(1), 35-54.

Yildirim, S., Narayanan, S. & Potamianos, A. (2011). Detecting emotional state of a child in a conversational computer game. In *Computer Speech & Language*, 25(1), 29-44. https://doi.org/10.1016/j.csl.2009.12.004

Zimmerman, E. (2009). Gaming literacy: Game design as a model for literacy in the twenty-first century. In Perron, B., Wolf, M.J. (Eds.), *The Video Game Theory Reader 2*. New York: Routledge, 23-31.

Appendix

Italian UES Short form and Core GL(Skill) scale

The validated Italian translations of the UES Short form and the videogaming skill scale (O'Brien et al., 2018; Rosenberg, 2011) are presented here together with the original items in a non-randomized order. With respect to the UES, participants were asked to fill in the questionnaire while keeping in mind their experience with the *Minecraft* gaming session. Then, the skill items were introduced by an alert indicating their more general nature. The following semantic tags were assigned to the Likert values: 1 = 'I totally disagree'; 2 = 'I disagree'; 3 = 'I neither disagree nor agree'; 4 = 'I agree'; 5 = 'I totally agree'.

FA 1. Mi sono immerso completamente in questa esperienza.
[I lost myself in this experience.]
FA 2. Il tempo è volato via mentre partecipavo al gioco.
[The time I spent using Application X just slipped away.]
FA 3. Sono stato preso da questa esperienza.
[I was absorbed in this experience.]
PU 1. (Reversed) Ho provato frustrazione a partecipare a questo gioco.
[I felt frustrated while using this Application X.]
PU 2. (Reversed) Ho trovato questo gioco disorientante.
[I found this Application X confusing to use.]
PU 3. (Reversed) Partecipare a questo gioco è stato faticoso.
[Using this Application X was taxing.]
AE 1. Questo gioco è stato accattivante.
[This Application X was attractive.]
AE 2. Questo gioco è stato esteticamente soddisfacente.
[This Application X was aesthetically appealing.]
AE 3. Questo gioco ha parlato ai miei sensi.
[This Application X appealed to my senses.]
RW 1. È valsa la pena di partecipare a questo gioco.
[Using Application X was worthwhile.]
RW 2. La mia esperienza è stata gratificante.
[My experience was rewarding.]
RW 3. Ho provato interesse per questa esperienza.
[I felt interested in this experience.]

CoreGL(Skill) 1. Sono abile a giocare ai videogiochi.
[I am skilled at playing video games.]
CoreGL(Skill) 2. Mi considero esperto delle mosse, delle tecniche e delle strategie nei videogiochi a cui gioco.
[I consider myself knowledgeable about moves, techniques, and strategies in/for the video games I play.]
CoreGL(Skill) 3. (Reversed) Conosco meno i videogiochi rispetto alle altre persone.
[I know less about games than most gamers.]
CoreGL(Skill) 4. Giocare ai videogiochi mi risulta facile.
[I find it easy to play video games.]
CoreGL(Skill) 5. So come fare ciò che voglio quando gioco ai videogiochi.
[I know how to do what I want when playing video games.]