Knowledge Tracing And Gamified Onboarding to Support (Language) Learning in a Platform With A Purpose for Collecting Linguistic Judgments

Chris Madge,[1] Doruk Kicikoglu,[1] Fatima Althani,[1]
Richard Bartle,[2] Jon Chamberlain,[2] Udo Kruschwitz,[3] Massimo Poesio[1]
[1] Queen Mary University Of London, [2] University of Essex, [3] University of Regensburg
{c.j.madge,d.kicikoglu,f.althani,m.poesio}@qmul.ac.uk, {rabartle,jchamb}@essex.ac.uk, Udo.Kruschwitz@ur.de

Abstract

Crowdsourcing tasks require training workers and testing their knowledge. This training requires, on the one hand, the development of appropriate tutorials; on the other, keeping track of the workers' understanding of the key concepts. In this paper, we present the novel Knowledge Tracing methods and onboarding methods developed for *LingoTowns*, a GWAP platform for collecting multiple types of linguistic judgments. *LingoTowns* is intended as a gamified language practice tool, and therefore requires methods for teaching linguistic concepts to its players and tracking their progress in absorbing such concepts.

Introduction

Crowdsourcing tasks require training workers and testing their knowledge (Kittur et al. 2011; Fortson et al. 2012; Dontcheva et al. 2014; Staffelbach et al. 2015; Mitra, Hutto, and Gilbert 2015). This training requires, on the one hand, the development of appropriate tutorials; on the other, keeping track of the workers' understanding of the key concepts.

Our area of research is GWAPs for language resources creation. The annotation of natural language data is a skilled task that generally requires training players/workers before they can achieve a thorough understanding of the necessary linguistic concepts (Lafourcade 2007; Guillaume, Fort, and Lefebvre 2016; Madge et al. 2019b; Kicikoglu et al. 2020). We present a GWAP platform for collecting multiple types of linguistic judgments that was developed to be used as a gamified language practice tool, and therefore was designed around teaching linguistic concepts to its players and tracking their progress in absorbing such concepts.

The first contribution of this work is that we propose the application in this setting, of a Knowledge Tracing (KT) method from the Intelligent Tutoring Systems literature–specifically, Bayesian Knowledge Tracing (BKT) (Corbett and Anderson 1994). Our second contribution is the application of such methods to a language practice platform designed to cover *multiple* linguistic interpretation levels, whereas most GWAPs for linguistic data are focused on one level of interpretation only–e.g., part-of-speech tagging (Madge et al. 2019a), lexical semantics (Vannella et al. 2014) or syntax (Fort, Guillaume, and Chastant 2014). Our platform was designed to train students / players starting from the simplest linguistic concepts all the way to more advanced ones. Finally, the design of the tutorial in GWAPs can greatly impact engagement, more so than entertainment games (Andersen et al. 2012). The third contribution of this paper concerns anonboarding design inspired by previous work (e.g. (Andersen et al. 2012; Poretski and Tang 2022; Thomsen et al. 2016)) to improve learnability in this crucial stage of the game.

Presented at the Works-in-Progress and Demonstrations track, AAAI HCOMP 2023.

Background

Progression in GWAPs Whilst to our knowledge, there is no GWAP work that has previously directly employed knowledge tracing, various works have employed employed progression algorithms and tutorials with a learning focus (Fort, Guillaume, and Chastant 2014; Madge et al. 2019a; Dumitrache et al. 2013). Probabilistic aggregation has been used to estimate task complexity, but this was focused on accuracy (Madge et al. 2019b) rather than learning. The system that bares most resemblance to our progression approach is *Quizz*, a gamified multiple-choice system to gather new facts (Ipeirotis and Gabrilovich 2014) and using a Markov Decision Process to model continued user interactions.

Knowledge Tracing in Intelligent Tutoring Systems In Computer Supported Intelligent Tutoring systems, the system aims to assess users knowledge to offer them questions and information that challenge their ability while optimally aiding their comprehension. The goal is to close the gap on individualised one-to-one human tuition in which the human can directly interact with the tutor (Corbett 2001). Similarly, in Computer Aided Testing there are often a large quantity of questions available to assess a users ability, and one of the goals of the systems are to offer a minimal number of interactions while best estimating a fair final score.

The BKT model is a Hidden Markov Model that pools over users and items, for each skill, to attempt to model the latent knowledge of a skill and the probability of a correct response given some observations (Corbett and Anderson 1994). We believe this could be suitable for addressing a number of challenges in GWAPs. For example, this could provide information regarding when a player has learned a skill without testing; the confidence in a label provided by an

Figure 1: *LingoTowns* - A Town Associated with a Document and its Locked Buildings

annotator; when to progress a player onto a new task; when and which educational material to show; and when to test against gold to discover player ability and when to annotate.

Onboarding in GWAPs

GWAPs can be more complex than entertainment games and can benefit from incorporating tutorials (Andersen et al. 2012). Tutorials are usually found in the first stage of a game, the onboarding phase. This is the stage where players are given a purpose to play (Chou 2019) and it is one of the most crucial stages when it comes to player engagement (Cheung, Zimmermann, and Nagappan 2014). This stage determines long term engagement as it teaches players the main concepts and mechanics of the game. They are vital to GWAPs, as players frequently face difficulties when interacting with the game (Miller and Cooper 2022). We incorporate other elements included in onboarding, such as narratives, that help provide players with a storyline, giving them a reason to play (Poretski and Tang 2022) and increase motivation (Lee et al. 2013).

LingoTowns

The context for the research in this paper is *LingoTowns* (Madge et al. 2022), a platform for collecting linguistic judgements at all levels of linguistic interpretation from language learners by providing gamified language practice exercises covering multiple aspects of grammar.

LingoTowns wraps GWAPs for the separate language levels in a procedurally generated, infinite virtual world environment. Each town in the virtual world is associated with a single document, and each building in that town (the cafeteria, the library, etc) is associated with an annotation activity/game for a specific language level. (Figure 1.) The players can only see play games associated with linguistic competencies for which they have been trained, but once they demonstrate their understanding of these competencies, they can undergo training for the next level and, if they demonstrate a sufficient levels of understanding, they can **unlock** new mini-games.

This design allows us to manage the players experience as they navigate between: different documents exhibiting different complexity of annotation instances to support the application of existing skills in increasingly complex settings; in game skills (e.g. discourse new, discourse old) to support development of existing skills; tasks/games (e.g. labelling part of speech, labelling coreference) to support teaching approaches such scaffolding and development of new skills.

This design also enables us to provide task specific onboarding and in-depth interactive tutorials to players.

Mini Games

There are currently three mini games in *LingoTowns* (more are planned for other aspects of grammatical competence). *CafeClicker* [1] is the simplest of the games, focusing on assigning parts-of-speech (e.g., noun, verb) to individual words. This first game serves as a comfortable sandbox (Gee 2004) to help the player / learner understand the process of annotation itself, and also, by learning about pronouns, nouns and other parts of speech, provides a foundation for their subsequent acquisition of the ability to learn (an annotate) syntactic constituents – in particular, noun phrases. *PhraseFarm* [2] is the second mini-game in the current linguistic pipeline; teaching/gathering annotations for, noun phrases. *Lingotorium* (Kicikoglu et al. 2019) [3] is the final mini-game, teaching and gathering annotations for anaphoric reference.

Adapting BKT for GWAPs

The progression mechanisms in *LingoTowns* and its mini-games hinted at in the previous Section are implemented using a form on BKT (the underlying implementation [4] based on the simulated annealing approach (Miller, Baker, and Rossi 2014)).

The core challenge of applying BKT in the GWAP context is that we are missing two key bits of information. The first is that, unlike the field of intelligent tutoring, we do not know the correct answer. In addition, on many occasions we may not know the correct skill. In this work, we assign the skill as being the context of the item (e.g. in the case of anaphoric reference: "Discourse New"; "Discourse Old"; "Property"; "Non-Referring", or in the case of parts-of-speech: noun; adjective; verb etc.).

As previously mentioned, in this setting we cannot know what the correct skill is with absolutely certainty. In ordinary BKT, one would estimate the correctness of the next answer given a certain knowledge over the skill required. For some coder, j and some skill/class k, the probability of being correct on the next round is ($P(S)$: slipping $P(G)$: guessing):

$$P(C_{t+1})_j^k = P(L_{t+1})_j^k(1 - P(S)^k) + (1 - P(L_{t+1})_j^k)p(G)^k$$

However, as we can not be certain of the skill, we utilise the probability of the correct skill from MPA ($P(c_i = k)$). Our estimate is the sum of the joint probability of being correct, for each skill, with some smoothing (smoothing factor $\lambda = 0.1$) to account for data sparsity. This makes the proba-

[1]https://cafeclicker.com

[2]https://phrasefarm.org

[3]https://lingotorium.com

[4]https://github.com/chrism-qmul/bkt

bility of coder j being correct on the next round:

$$P(C_{t+1})_j = \sum_{k}^{K} P(C_{t+1})_j^k P(c_i = k)(1 - \lambda) + \frac{\lambda}{K}$$

Player Training as Grammatical Concepts Learning in *LingoTowns*

Through *LingoTowns*, we offer a gamified curriculum learning approach to teaching the players grammatical concepts, using skill chains (discussed in following section). By utilising previously learned related skills, we switch between various games to promote the acquisition of new complex abilities (for example, learning about nouns scaffolds noun-phrase understanding).

Our *LingoTowns* progression allows players to practice skills on increasingly difficult documents, so that the player benefits from becoming accustomed to applying a skill on basic linguistic phenomena for that annotation task, before applying that skill in more difficult contexts, offering the player to practice in "cycles of expertise" (Gee 2004) (i.e. skills within tasks; tasks within documents). We can also manage their annotation interactions from the lifetime of a skill, minimising the requirement to apply it once it reaches the "burn out" stage.

Breaking down the design into a set of skill atoms gives a systematic approach to identifying the individual skills or "knowledge components", with which we use our bespoke approach to identify how a player is progressing. Applying KT allows us to estimate when the player has learned a skill, and is ready to be introduced to a new challenge. Our skill chain (visualised in Figure 2) is designed to encode an order in which grammatical concepts are learned that reflects, on the one hand, what is a natural order in which these concepts are acquired; on the other, the dependencies that exist between the three types of annotation we gather (i.e. anaphora links noun phrases; noun-phrases typically contain some form of noun).

Onboarding in *LingoTowns*

The onboarding phase in *LingoTowns* begins by introducing players to the story. Narratives provide players with a purpose to play the game (Poretski and Tang 2022). To improve learnability we have adopted the multimedia learning principles set out by Mayet et al. (Mayer and Moreno 2002). For example, following the personalisation principle, we present descriptions and narrative in a conversational style, with characters acting as advisors (Poretski and Tang 2022). When the text is shown, it is coupled with animation, in line with the multiple representation principle.

In each of the mini-games the player is walked through annotating their first sentence interactively using the same characters, highlights of key user interface components and discussion of game-play mechanics. The games also feature sections in which the more eager players can go beyond the tutorial to learn more.

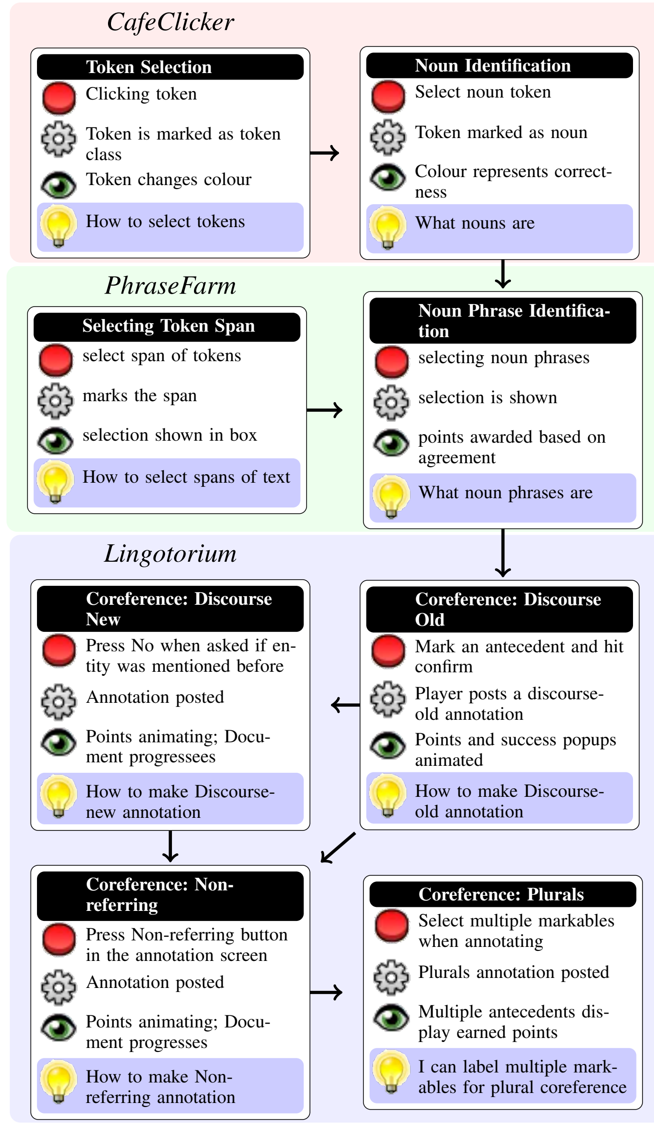

Figure 2: Skill Chain in *LingoTowns*

Conclusion

We believe learning to be a key component of skilled GWAPs, and one that is not considered as an objective in the current progression models used in GWAPs. The application of these methods could add a valuable source of information to progression and learning in GWAPs, in turn further increasing engagement and player enjoyment.

In this work, we present the approach to language learning followed in the *LingoTowns* platform and its sub-games. We make a case for using KT methods in Games With A Purpose, and demonstrate a bespoke implementation of these methods that we use to track skill chain-based progression between games and within each game.

Future work will test this approach, examining the impact on player learning, long term accuracy, engagement and other factors.

References

Andersen, E.; O'rourke, E.; Liu, Y.-E.; Snider, R.; Lowdermilk, J.; Truong, D.; Cooper, S.; and Popovic, Z. 2012. The impact of tutorials on games of varying complexity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 59–68.

Cheung, G. K.; Zimmermann, T.; and Nagappan, N. 2014. The first hour experience: how the initial play can engage (or lose) new players. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play*, CHI PLAY '14, 57–66. New York, NY, USA: Association for Computing Machinery.

Chou, Y.-K. 2019. *Actionable Gamification: Beyond Points, Badges, and Leaderboards*. Packt Publishing Ltd.

Corbett, A. 2001. Cognitive computer tutors: Solving the two-sigma problem. In *International Conference on User Modeling*, 137–147. Springer.

Corbett, A. T.; and Anderson, J. R. 1994. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4(4): 253–278.

Dontcheva, M.; Morris, R. R.; Brandt, J. R.; and Gerber, E. M. 2014. Combining crowdsourcing and learning to improve engagement and performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3379–3388.

Dumitrache, A.; Aroyo, L.; Welty, C.; Sips, R.-J.; and Levas, A. 2013. Dr. Detective": combining gamification techniques and crowdsourcing to create a gold standard in medical text. In *Proceedings of the 1st International Workshop on Crowdsourcing the Semantic Web (CrowdSem 2013), 12th International Semantic Web Conference*. Citeseer.

Fort, K.; Guillaume, B.; and Chastant, H. 2014. Creating Zombilingo, a Game With A Purpose for dependency syntax annotation. In *Proceedings of the First International Workshop on Gamification for Information Retrieval*, 2–6.

Fortson, L.; Masters, K.; Nichol, R.; Edmondson, E.; Lintott, C.; Raddick, J.; and Wallin, J. 2012. Galaxy zoo. *Advances in machine learning and data mining for astronomy*, 2012: 213–236.

Gee, J. P. 2004. Learning by design: Games as learning machines. *Interactive educational multimedia: IEM*, 15–23.

Guillaume, B.; Fort, K.; and Lefebvre, N. 2016. Crowdsourcing complex language resources: Playing to annotate dependency syntax. In *International Conference on Computational Linguistics (COLING)*.

Ipeirotis, P. G.; and Gabrilovich, E. 2014. Quizz: targeted crowdsourcing with a billion (potential) users. In *Proceedings of the 23rd international conference on World wide web*, 143–154.

Kicikoglu, O. D.; Bartle, R.; Chamberlain, J.; Paun, S.; and Poesio, M. 2020. Aggregation Driven Progression System for GWAPs. In *Workshop on Games and Natural Language Processing*, 79–84.

Kicikoglu, O. D.; Bartle, R.; Chamberlain, J.; and Poesio, M. 2019. Wormingo: a 'true gamification'approach to anaphoric annotation. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*, 1–7.

Kittur, A.; Smus, B.; Khamkar, S.; and Kraut, R. E. 2011. Crowdforge: Crowdsourcing complex work. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, 43–52.

Lafourcade, M. 2007. Making people play for Lexical Acquisition with the JeuxDeMots prototype. In *SNLP'07: 7th international symposium on natural language processing*, 7.

Lee, T. Y.; Dugan, C.; Geyer, W.; Ratchford, T.; Rasmussen, J.; Shami, N. S.; and Lupushor, S. 2013. Experiments on motivational feedback for crowdsourced workers. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 7, 341–350.

Madge, C.; Bartle, R.; Chamberlain, J.; Kruschwitz, U.; and Poesio, M. 2019a. Incremental game mechanics applied to text annotation. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 545–558.

Madge, C.; Brightmore, J.; Kicikoglu, D.; Althani, F.; Bartle, R.; Chamberlain, J.; Kruschwitz, U.; and Poesio, M. 2022. LingoTowns: A Virtual World For Natural Language Annotation and Language Learning. In *Extended Abstracts of the 2022 Annual Symposium on Computer-Human Interaction in Play*, 57–62.

Madge, C.; Yu, J.; Chamberlain, J.; Kruschwitz, U.; Paun, S.; and Poesio, M. 2019b. Progression in a language annotation game with a purpose. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 7, 77–85.

Mayer, R. E.; and Moreno, R. 2002. Animation as an aid to multimedia learning. *Educational psychology review*, 14: 87–99.

Miller, J. A.; and Cooper, S. 2022. Barriers to expertise in citizen science games. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–25.

Miller, W. L.; Baker, R. S.; and Rossi, L. M. 2014. Unifying computer-based assessment across conceptual instruction, problem-solving, and digital games. *Technology, Knowledge and Learning*, 19(1): 165–181.

Mitra, T.; Hutto, C. J.; and Gilbert, E. 2015. Comparing person-and process-centric strategies for obtaining quality data on amazon mechanical turk. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 1345–1354.

Poretski, L.; and Tang, A. 2022. Press A to Jump: Design Strategies for Video Game Learnability. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–26.

Staffelbach, M.; Sempolinski, P.; Kijewski-Correa, T.; Thain, D.; Wei, D.; Kareem, A.; and Madey, G. 2015. Lessons learned from crowdsourcing complex engineering tasks. *PloS one*, 10(9): e0134978.

Thomsen, L. E.; Petersen, F. W.; Drachen, A.; and Mirza-Babaei, P. 2016. Identifying Onboarding Heuristics for Free-to-Play Mobile Games: A Mixed Methods Approach. In

Wallner, G.; Kriglstein, S.; Hlavacs, H.; Malaka, R.; Lugmayr, A.; and Yang, H.-S., eds., *Entertainment Computing - ICEC 2016*, 241–246. Cham: Springer International Publishing. ISBN 978-3-319-46100-7.

Vannella, D.; Jurgens, D.; Scarfini, D.; Toscani, D.; and Navigli, R. 2014. Validating and extending semantic knowledge bases using video games with a purpose. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1294–1304.