A dialogue-based software architecture for gamified logopedic discrimination tests.

Antonio Origlia
Dept. of Information Engineering
University of Padova
Via G. Gradenigo 6/b
35131 Padova, Italia
antori@gmail.com

Piero Cosi
Istituto di Scienze e Tecnologie della Cognizione (ISTC)
Via Martiri della Libertà 2
35137 Padova, Italia
piero.cosi@pd.istc.cnr.it

Antonio Rodà
Dept. of Information Engineering
University of Padova
Via G. Gradenigo 6/b
35131 Padova, Italia
roda@dei.unipd.it

Claudio Zmarich
Istituto di Scienze e Tecnologie della Cognizione (ISTC)
Via Martiri della Libertà 2
35137 Padova, Italia
claudio.zmarich@cnr.it

ABSTRACT

In this work we describe the current stage of development of a software architecture designed to present logopedic tests to pre-school children in the form of gamified tasks. We interpret the problem of administering these tests as a dialogue model using probabilistic rules to generate customised tests on the basis of the child's performance. In the proposed architecture, the dialogue system controls a gaming setup composed of a virtual agent and a robotic companion that needs to be taught how to talk. This learning-by-teaching approach is used to camouflage a phonemes discrimination test that has the added value of being generated at runtime on the basis of the child's performance. We will describe the architectural components involved in the architecture and and we will present a probabilistic model describing a logopedic test. The selected stimuli are then administered by generating instructions for the virtual agents involved in the game.

Author Keywords

Guides; instructions; author's kit; conference publications; keywords should be separated by a semi-colon.
Optional section to be included in your final version, but strongly encouraged.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Paste the appropriate copyright statement here. ACM now supports three different copyright statements:
• ACM copyright: ACM holds the copyright on the work. This is the historical approach.
• License: The author(s) retain copyright, but ACM receives an exclusive publication license.
• Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.
This text field is large enough to hold the appropriate release statement assuming it is single spaced.
Every submission will be assigned their own unique DOI string to be included here.

See: `http://www.acm.org/about/class/1998/` for more information and the full list of ACM classifiers and descriptors. Optional section to be included in your final version, but strongly encouraged. On the submission page only the classifiers letter-number combination will need to be entered.

INTRODUCTION

[Qualche parola da Claudio sui test di discriminazione?]

[Qualche parola da Piero sull'ILT?]

Healthy contact with language, in the first years of life, consists of a playful activity where parents and infants engage *protoconversations* made of rhythmical and musical content. This manifests the emotional regulation of *primary intersubjectivity* [17], where interaction with the caregiver, either reciprocally directed or mediating access to objects of interest for the infant, manifests the typical playfulness often observed in mammals. At 9 months, *secondary intersubjectivity* arises [21] and the baby's interest moves onto sharing the ways companions use objects as she starts to interact with the material world in a more informed way. The caregivers' language also shifts, in this phase, from questions and rhetorical comments to instructions and informative comments to support the baby's interest in participating to a task [8]. This is "[. . .] the start of cultural information transfer between generations" [18, p. 74]. Playful behaviour adapts to new roles as the child grows older but always stays in the background, motivating access to cultural information, reinforcing memory and supporting the creation of meaning [20, 13]. Language development strongly depends on intersubjective experiences: from the effective engagement of minds and bodies depends cultural learning [4]. Although humans appear to be born with a natural disposition towards cultural learning [19], successful acquisition of cultural skills depend on the interaction quality, especially considering social feedback. Perceived affection and playfulness by the infant towards the parents helps establish a mutually teasing situation [13] that focuses attention on rituals that may later become skills [6].

Given the social nature of cultural transfer, it is not enough to expose children to new words without providing an adequate *context* to them. Engaging and meaningful activities are especially important to attract interest in the children and *show* them how words can provide the natural pleasure that comes with gaining competence in interacting with their loved ones and with peers. Storytelling has been demonstrated to be a powerful mean to accomplish this as children are born with "[...] an abundant and early armament of narrative tools" [1, p. 90]. Through storytelling, children acquire skills related to the so called *emergent literacy* [15], which is a necessary prerequisite to mastering reading and writing. These skills cover metalinguistic awareness, cohesion and reference in oral communication and the capability of making one's own intentions known to others. Emergent literacy capabilities "[...] are acquired first in language play and in storytelling. Many of them are acquired in the context of childrens interactions with peers, in early play contexts." [2, p. 76]. Once again, the importance of social context and playful interaction is clearly highlighted concerning the acquisition of literacy skills. Wordplay for children appears to be based on matching and substituting words on the basis of their sound rather than their meaning as they appear to [2, p. 78] "[...] derive tremendous pleasure from rhyming words ("you silly"; "no, you pilly") or words that sound similar (adult: "Indians lived in a teepee"; child: "pee-pee!")". In order to exploit the children's way of *having fun* with words, it is necessary to go beyond the concept of *correct* and *wrong* words when designing playful activities. It is through this kind of judgement-free exploration of words aimed at sharing playful intents that they can develop their metalinguistic awareness that language is produced for other people as they attempt to involve them into play. In order to become meaningful and precious for children, teaching activities need them to have a basis of experiences showing language as a tool to provide pleasure both to the self and to companions in social activities.

Other than supporting the acquisition of literacy skills for children in general, social and emotional awareness is of extreme importance when designing systems that aim at providing support to children suffering of dyslexia. Past studies highlighted how the emotional concept of *happiness* is associated with the concept of possessing *good reading ability* by children suffering from dyslexia in a much stronger way than children without dyslexia [16]. Subjects suffering from dyslexia report to feel "[...] disappointed, frustrated, ashamed, fed up, sad, depressed, angry and embarrassed by their difficulties" [14, p. 129]. Consequences of the load of negative emotions caused by dyslexia has been describes as *scars* in [7]. The importance of the affective component in children with reading and writing difficulties suggests that artificial agents may appeal to these emotions to establish an empathic connection with the children and gently support them in rebuilding their self-confidence.

Acquiring expertise in the subtleties of the use of language results in social and personal value improves self-confidence and the quality of interaction with others. Controlling the higher aspects of language, however, requires children to complete and refine their phonological and phonetic competence. Working on such details can easily become frustrating, tiresome and even pointless, from the child's point of view, if playfulness is substituted with coercion and unmotivated requests to produce efforts. A carefully designed learning environment should take into account the strong emotional component involved in cultural transfer as it implicates that spontaneous interest in accessing literacy heritage emerges only when this is functional to access fun, pleasant experiences. A child's motives to improve her literacy competence should come *from the inside* by making literacy an object of interest and pleasure.

SYSTEM ARCHITECTURE

[FARE LA NUOVA IMMAGINE]

LINGUISTIC KNOWLEDGE BASE

With the advent of the Big Data and, in particular, with the increasing availability of Linked Open Data, the need to establish a representation format suitable for dynamic, rapidly changing and interconnected *objects* arose. RDF represents the most widely used solution to this need and has been adopted to implement the most widely known repositories of linked knowledge available today. RDF is based on the concept of triple stores: every *statement* about a specific domain is represented by a *subject*, a *predicate* and an *object*. From a conceptual point of view, an RDF statement represents a link, whose semantic value is described by the predicate, between two objects. The RDF standard has been developed in the last decade mainly to support semantic web applications and, because of this, it is heavily focused on the concept of Uniform Resource Identifier (URI), later generalised to International Resource Identifier (IRI). In RDF, every concept is represented by an IRI and there is no explicit difference between items and relationships, which are only distinguished by their *syntactic* role in the statement. Although RDF is agnostic about the semantic meaning of IRIs, these can be unequivocally interpreted if a source vocabulary is specified. For example the IRI 'http://dbpedia.org/resource/Name' specifies that the item *Name* is supposed to be interpreted in the way specified by DBpedia. Whenever this property is used while specifying the DBpedia vocabulary, all clients accessing the item are informed about its meaning. While the set of RDF triple stores defines a graph of relationships, the representation is edge-centred in the sense that there is no way to specify item or relationship properties in any other way than by establishing new edges.

Graph databases, on the other hand, are node-centred in the sense that the main data container are the graph nodes, with relationships being separated objects describing how these nodes are related to each other. This conceptual separation between nodes and relationships is one of the main differences between RDF and graph databases. Graph databases also explicitly distinguish *properties* from relationships. While in RDF an item is linked to its properties by the use of relationships, graph databases explicitly represent properties as part of either nodes or relationships. Another fundamental difference between RDF and graph databases is that, in RDF, it is not possible to specify properties for a specific relationship between two nodes.

From the data access point of view, graph databases and RDF both have their strengths and weaknesses so choosing one over the other mainly depends on the intended use. Graph databases are designed to be very efficient in graph traversing and path finding. RDF has, instead, powerful support to relationship inference through the application of logical rules.

Neo4J [22] is an open source graph database manager that has been developed over the last sixteen years and has been applied to a high number of tasks related to data representation [3], exploration [5] and visualisation [9]. It can be deployed in server mode and be queried over a specific port using HTTP or Bolt protocols, or it can be embedded in Java applications through dedicated APIs. In Neo4J, nodes and relationships may be assigned *labels*, which describe the type of the object they are associated to. In this work, labels are mainly used to represent morpho-syntactic characteristics of words and the nature of the relationships among nodes. Nodes and relationships may have *properties*, which are used here to store the details of each single node or relationship. Labels and properties are the main way used by Neo4J to filter data and retrieve answers to user queries.

In this work, we use the MultiWordNet-Extended (MWN-E) dataset [11], as the knowledge base to control the decision process for the logopedic discrimination test. The MWN-E dataset is based on the MultiWordNet dataset [12] and extended by introducing morpho-syntactic data (e.g. gender, number…), derived forms (e.g. plurals, conjugations…) and SAMPA pronunciations. Also, phonological neighbourhoods are computed and are of particular interest for this work. A word A is defined to be a phonological neighbour of the word B if it is possible to obtain B by altering the transcription of A using exactly one Insertion/Deletion/Substitution operation. Phonological neighbourhoods are represented by establishing relationships of type HAS_PHONOLOGICAL_NEIGHBOUR between two words if their Minimum Edit Distance equals 1. This kind of relationship has a *distance* property that, in these cases, is set to 1. Relationships of type HAS_PHONOLOGICAL_NEIGHBOUR are also established between words that have the same pronunciation but have different written forms. In this case the value of the *distance* property is set to 0. Other than the data included in the version presented in [11], the MWN-E version used in this work also contains frequency data for the terms in the vocabulary extracted from the *Primo Vocabolario del Bambino* (First vocabulary of children) **[RIFERIMENTO]** and from the Italian Wikipedia[1]. Currently, MWN-E consists of 292282 nodes containing 1442579 properties. 943174 relationships among these nodes are found, phonological neighbourhood relationships at distance 1 representing the majority.

It is important to note that, although the phonological neighbourhood relationship is oriented, like all the connections in Neo4J, it is bidirectional by nature and should be queried using orientation independent syntax in extracting data from graphs using the Neo4J querying language: Cypher. Cypher is designed to be a *declarative* language that highlights patterns structure by using an SQL inspired *ascii-art syntax*. A brief overview of the syntactic elements of Cypher queries is given here to help understanding the example queries presented in this paper. The reader is referred to the online Cypher manual[2] for a more detailed presentation of Cypher. As in graphical representations of graphs nodes are usually represented by circles, in Cypher nodes are represented by round brackets. For example, the query "MATCH (n:VERB) RETURN n" returns all the nodes of the graph labelled as verbs. In the same way, since relationships are usually represented by labelled arrows in graph schemas, relationships between nodes are described by using ASCII *arrows*, too. The query "MATCH (m)-[:DERIVES_FROM]-¿(:VERB word: 'essere') RETURN m" returns all the nodes that contain a term that derives from the *essere* (to be) verb. The SQL-like WHERE clause may also be used to filter results using boolean logic.

OPENDIAL

Opendial [10] is a dialogue management framework based on probabilistic rules which aims to merge the best of rule-based and probabilistic dialogue management. In cases where a good amount of previous knowledge about the domain is possessed by the dialogue designer with specific needs of fine-tuning rules, the rule based approach can be integrated with probability and utility-based reasoning to fine tune the system's response. Probabilistic rules, in Opendial, are used to setup and update a bayesian network consisting of variables representing the dialogue state. Depending on this, the dialogue manager selects the most probable user action given a set of, possibly inaccurate, inputs. Using a set of utility functions provided by the dialogue designer, the manager computes the most *useful* system reaction, possibly generating natural language responses or executing actions. In Opendial, it is possible to apply *a priori* estimates on future values of state variables. The probability distributions providing *a priori* estimates can be updated, using bayesian inference, after the actual observation arrives to dynamically improve the model. In Opendial, dialogue domains are described in an XML format specifically designed for the dialogue system. This is composed of a set of models triggered by variable updates, of rules, included in the model and changing the dialogue state according to the dialogue designer description. Opendial supports unification in its dialogue specification language so that variables can be included to obtain generic rules. In the example shown in Figure 1, a part of the model that identifies opposing traits given two phonemes is presented. The condition for the considered rule to fire is that the two phonemes in the *opposition* variable are not the same one. If this is the case, a user defined *HasTraits* function is used to determine if the two phonemes have the *sonorant* trait. Then, probability of the set of opposing traits to contain the *sonorant* trait is equal to the XOR of the result obtained by applying the *HasTraits* function on the considered phonemes.

[1]Data extracted from the 20/04/2017 Wikipedia.it dump

[2]https://neo4j.com/developer/cypher-query-language

```
<rule>
    <case>
        <condition>
            <if var="opposition" relation="contains" value="{X};{Y}"/>
            <not>
                <if var="opposition" relation= "contains" value= "{X};{X}"/>
            </not>
        </condition>
        <effect prob=" (HasTraits({X},Sonorant) -
                        HasTraits({Y},Sonorant)) *
                        (HasTraits({X},Sonorant) -
                        HasTraits({Y},Sonorant))">
            <set var="OpposedTraits" value="{OpposedTraits}+Sonorant"/>
        </effect>
    </case>
</rule>
```

Figure 1. Example rule to check whether two phonemes have the Sonorant train in opposition.

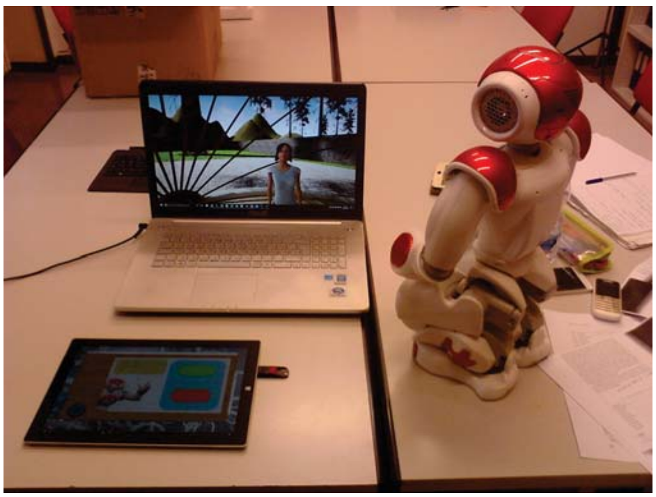

Figure 2. The experimental setup. [RIFARE BENE L'IMMAGINE]

Opendial can also be extended with Java-based plugins. In our case, we developed a set of plugins to connect the dialogue system to the Neo4J database and to the remote actors providing the user interface. The system makes use of the prediction and feedback mechanism provided by Opendial to build the probability distributions describing the likelihood of a subject to discriminate a specific trait and to select the next optimal stimulus to improve the user model given previous answers, thus obtaining an adaptive test.

INTERFACE

2

3

4

3

4

5

CONCLUSION

[3]www.softbank.jp/en/corp/group/sbr/
[4]www.unrealengine.com
[5]www.mivoq.it

Figure 3. The *teacher* virtual character, Ellie.

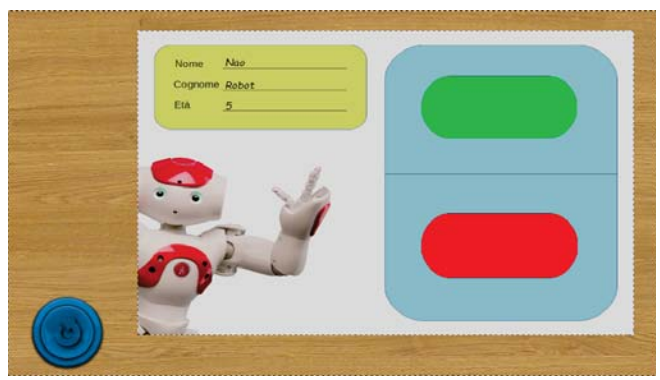

Figure 4. The tablet interface. The evaluation card enters the scene after the opposed words have been presented. The child gives feedback by touching the red or green areas to put a *stamp* to evaluate Nao's performance. A repeat button is also present to allow the child to listen to the opposed words again. This is allowed only once in the test.

ACKNOWLEDGMENTS

Antonio Origlia's work is supported by Veneto Region and European Social Fund (grant C92C16000250006).

REFERENCES

1. Bruner, J. S. *Acts of meaning*, vol. 3. Harvard University Press, 1990.
2. Cassell, J. Towards a model of technology and literacy development: Story listening systems. *Journal of Applied Developmental Psychology 25*, 1 (2004), 75–105.
3. Dietze, F., Karoff, J., Valdez, A. C., Ziefle, M., Greven, C., and Schroeder, U. An open-source object-graph-mapping framework for neo4j and scala: Renesca. In *International Conference on Availability, Reliability, and Security*, Springer (2016), 204–218.
4. Donald, M. *A mind so rare: The evolution of human consciousness*. WW Norton & Company, 2001.
5. Drakopoulos, G., Kanavos, A., Makris, C., and Megalooikonomou, V. On converting community detection algorithms for fuzzy graphs in neo4j. In *Proceedings of the 5th International Workshop on Combinations of Intelligent Methods and Applications, CIMA* (2015).
6. Eckerdal, P., and Merker, B. Musicand the action songin infant development: An interpretation. *Communicative musicality: Exploring the basis of human companionship* (2009), 241–262.
7. Edwards, J. The scars of dyslexia. *Educational Psychology in Practice 11*, 2 (1995), 55.
8. Halliday, M. A. K. *Learning How to Mean–Explorations in the Development of Language*. ERIC, 1975.
9. Jiménez, P., Diez, J. V., and Ordieres-Mere, J. Hoshin kanri visualization with neo4j. empowering leaders to operationalize lean structural networks. *Procedia CIRP 55* (2016), 284–289.
10. Lison, P., and Kennington, C. Opendial: A toolkit for developing spoken dialogue systems with probabilistic rules. *ACL 2016* (2016), 67.
11. Origlia, A., P. G. C. F. Mwn-e: a graph database to merge morpho-syntactic andphonological data for italian. In *Proceedings of Subsidia* (2017).
12. Pianta, E., Bentivogli, L., and Girardi, C. Developing an aligned multilingual database. In *Proc. of the 1st International Conference on Global WordNet* (2002).
13. Reddy, V. *How infants know minds*. Harvard University Press, 2008.
14. Riddick, B. *Living with dyslexia: The social and emotional consequences of specific learning difficulties/disabilities*. Routledge, 2009.
15. Teale, W. H., and Sulzby, E. *Emergent Literacy: Writing and Reading. Writing Research: Multidisciplinary Inquiries into the Nature of Writing Series.* ERIC, 1986.
16. Thomson, M. E., and Hartley, G. M. Self-concept in dyslexic children. *Academic Therapy* (1980).
17. Trevarthen, C. Communication and cooperation in early infancy: A description of primary intersubjectivity. *Before speech: The beginning of interpersonal communication* (1979), 321–347.
18. Trevarthen, C. The functions of emotion in infancy. In *The healing power of emotion: Affective neuroscience, development & clinical practice (Norton Series on Interpersonal Neurobiology)*, D. Fosha and S. M. F. Siegel, D. J., Eds. WW Norton & Company, 2009, 55–85.
19. Trevarthen, C., and Aitken, K. Regulation of brain development and age-related changes in infants. *Motives: The Developmental Function of Regressive Periods.In M. Heimann (ed.) Regression Periods in Human Infancy. Mahwah, NJ: Erlbaum* (2003), 107–184.
20. Trevarthen, C., and Aitken, K. J. Infant intersubjectivity: Research, theory, and clinical applications. *Journal of child psychology and psychiatry 42*, 1 (2001), 3–48.
21. Trevarthen, C., Hubley, P., et al. Secondary intersubjectivity: Confidence, confiding and acts of meaning in the first year. *Action, gesture and symbol: The emergence of language* (1978), 183–229.
22. Webber, J. A programmatic introduction to neo4j. In *Proceedings of the 3rd annual conference on Systems, programming, and applications: software for humanity*, ACM (2012), 217–218.