University of Zurich[UZH]

Institut für Computerlinguistik

Cost-effectiveness of Games with a Purpose

Module «Crowdsourcing für Sprachtechnologie»

Mathias Müller

16. Mai 2019

Defining the object of investigation

Interested in the cost-effectiveness (CE) of software that

- is a game with a purpose (GWAP)
- makes users produce linguistically enriched data

How many NLP GWAP are there?

Defining the object of investigation

Interested in the cost-effectiveness (CE) of software that

- is a game with a purpose (GWAP)
- makes users produce linguistically enriched data

How many NLP GWAP are there?

A comprehensive list of GWAP in published literature

1001 Paraphrases (Chklovski, 2005)
FaCtory (Chamberlain et al., 2013)
Categorilla (Vickrey et al., 2008)
Categodzilla (Vickrey et al., 2008)
PlayCoref (Hladká et al., 2009)
PackPlay (Chamberlain et al., 2013)
GIVE (Chamberlain et al., 2013)
OntoGame (Siorpaes and Hepp, 2008)
Knowledge Towers (Vannella et al., 2014)
Puzzle Racer (Jurgens and Navigli, 2014)
JDM-pt (Mangeot and Ramisch, 2012)
Metropol Italia (Bry et al., 2013)
Kaboom! (Jurgens and Navigli, 2014)
An (unfortunately) unnamed game
in Pearl and Steyvers (2010)

LEARNER (Chamberlain et al., 2013)
Verbosity (von Ahn et al., 2006)
Free Association (Vickrey et al., 2008)
Phrase Detectives (Poesio et al., 2013)
PhraTris (Attardi, 2010)
Sentiment Quiz (Scharl et al., 2012)
JeuxDeMots (Lafourcade and Joubert, 2008)
Infection (Vannella et al., 2014)
SuchGame (Chamberlain et al., 2013)
ColorIt (Lafourcade et al., 2014)
OnToGalaxy (Krause et al., 2010)
Doodling (Kumaran et al., 2014)
Wordrobe (Venhuizen et al., 2013)

An (allegedly) complete list of GWAP that

- are documented in published literature
- generate a linguistic resource

A more readable list of GWAP for NLP

1001 Paraphrases	FaCtory	Categorilla
Free Association	PlayCoref	PackPlay
GIVE	OntoGame	Knowledge Towers
Puzzle Racer	JDM-pt	Metropol Italia
Kaboom!	LEARNER	Verbosity
Phrase Detectives	PhraTris	Sentiment Quiz
JeuxDeMots	Infection	SuchGame
ColorIt	OnToGalaxy	Wordrobe

Omitted: *Doodling* and the unnamed game from Pearl and Steyvers (2010)

Games that are online and can be played

1001 Paraphrases	FaCtory	Categorilla
Free Association	PlayCoref	PackPlay
GIVE	OntoGame	Knowledge Towers
Puzzle Racer	JDM-pt	Metropol Italia
Kaboom!	LEARNER	Verbosity
Phrase Detectives	PhraTris	Sentiment Quiz
JeuxDeMots	Infection	SuchGame
ColorIt	OnToGalaxy	Wordrobe

- online and playable
- difficult to access
- offline

Games that are online and can be played

At least half of the games are offline

- some never got past an experimental alpha phase
- ephemeral, short online presence

Being offline and half-baked is rather detrimental to cost-effectiveness

“Cost” mentioned in published literature

1001 Paraphrases	FaCtory	Categorilla
Free Association	PlayCoref	PackPlay
GIVE	OntoGame	Knowledge Towers
Puzzle Racer	JDM-pt	Metropol Italia
Kaboom!	LEARNER	Verbosity
Phrase Detectives	PhraTris	Sentiment Quiz
JeuxDeMots	Infection	SuchGame
ColorIt	OnToGalaxy	Wordrobe
Doodling	[unnamed game]	

- cost mentioned
- personal communication with author
- no mention

“Cost” mentioned in published literature *and* online

1001 Paraphrases	FaCtory	Categorilla
Free Association	PlayCoref	PackPlay
GIVE	OntoGame	**Knowledge Towers**
Puzzle Racer	JDM-pt	Metropol Italia
Kaboom!	LEARNER	Verbosity
Phrase Detectives	PhraTris	Sentiment Quiz
JeuxDeMots	**Infection**	SuchGame
ColorIt	OnToGalaxy	Wordrobe
Doodling	[unnamed game]	

- cost mentioned, online and playable

“Cost” mentioned in published literature *and* online

Two important observations:

- interestingly, almost all games where cost was mentioned are still online
- in general, neither being online for a long time nor cost are of importance in the literature

How to judge the CE of GWAP when it is only rarely mentioned in the literature?

Gather information about CE of GWAP

Ask researchers directly about cost:

I am wondering whether you have any rough estimate of ***how much money was used to create the game*** *(salary of all people involved, licensing of third-party software, server costs, hardware etc.).*
Jon Chamberlain and colleagues have published similar data for "Phrase Detectives" and I would like to be able to ***compare them****.*

Gather information about CE of GWAP

Ask researchers directly about the value of enriched data:

The users playing your game have generated linguistic data. Do you have a rough estimate of ***how much money it would have cost to obtain the very same data by traditional means*** *(i.e. paying experts for annotation tasks)?*

Gather information about CE of GWAP

Ask researchers directly about the CE:

In the specific case of your game, ***is developing a GWAP a cost-effective approach*** *that "pays off"? Do the gains from your game outweigh the costs?*

Gather information about CE of GWAP

Questions sent to 10 main authors

- Some people went offline along with their games
- Only 5 of them responded, 2 of which are the developers of GWAP that do mention cost

Respondent	**GWAP**
Arno Scharl	Sentiment Quiz
Markus Krause	OnToGalaxy
Mathieu Lafourcade	JeuxDeMots
Johan Bos	Wordrobe
Jon Chamberlain	Phrase Detectives

Some responses

Prof Arno Scharl:

Unfortunately we have collected ***no data whatsoever regarding the economic aspects*** *that you are investigating.*

Some responses

Prof Arno Scharl:

Unfortunately we have collected ***no data whatsoever regarding the economic aspects*** *that you are investigating.*

Prof Johan Bos:

We are ***actually now looking into*** *comparing Wordrobe with other crowd-sourcing methods (such as Crowdflower). Therefore I can't disclose any concrete figures right now.*

Estimating Cost-effectiveness

CE of GWAP as compared to Crowd-sourcing

What do we mean by “cost-effectiveness” of an NLP GWAP?

- purpose: get as much annotated data as possible, given the research budget
- cost-effective if the data it provided **could not have been obtained with less money**

Usually, GWAP are compared against Crowd-sourcing (CS) approaches.

CE of GWAP as compared to Crowd-sourcing

All of them compare against crowd-sourcing:

GWAP	**Cheapest**	**Fastest**
Doodling	GWAP	GWAP
TKT	GWAP	CS
Infection	GWAP	CS
Puzzle Racer	GWAP	CS
JeuxDeMots	GWAP	CS
PD	GWAP	CS

TKT = The Knowledge Towers, CS = Crowd-sourcing, PD = Phrase Detectives

CE of GWAP as compared to Crowd-sourcing

GWAP	Cheapest	Fastest	Formula for Success
Doodling	GWAP	GWAP	short-circuiting player attraction
TKT	GWAP	CS	students did all the work
Infection	GWAP	CS	students did all the work
Puzzle Racer	GWAP	CS	students did all the work
JeuxDeMots	GWAP	CS	longevity, attractiveness
PD	GWAP	CS	longevity, attractiveness

TKT = The Knowledge Towers, CS = Crowd-sourcing, PD = Phrase Detectives

CE of GWAP as compared to Crowd-sourcing

GWAP	Cheapest	Fastest	Formula for Success
Doodling	GWAP	GWAP	short-circuiting player attraction
TKT	GWAP	CS	students did all the work
Infection	GWAP	CS	students did all the work
Puzzle Racer	GWAP	CS	students did all the work
JeuxDeMots	GWAP	CS	longevity, attractiveness
PD	GWAP	CS	longevity, attractiveness

TKT = The Knowledge Towers, CS = Crowd-sourcing, PD = Phrase Detectives

- Key insight: GWAP is often cost-effective, but slow

The Cost-effectiveness of JeuxDeMots

Finally getting down to numbers

- size of lexical network: 20 million relations between 500 thousand terms
- speed of 1 hypothetical linguist: 6 relations per minute
- cost of said linguist: 30 CHF per hour

Comparing against traditional annotation:

$$20000000 \ / \ 6 = 3300000 \ min = 55555 \ h * 30 \ CHF = \mathbf{1666650 \ CHF}$$

The Cost-effectiveness of JeuxDeMots

Finally getting down to numbers

- size of lexical network: 20 million relations between 500 thousand terms
- cost of annotating a relation by a turker: 0.1 CHF

Comparing against crowd-sourcing:

$$20000000 * 0.1 = \mathbf{2000000\ CHF}$$

The Cost-effectiveness of JeuxDeMots

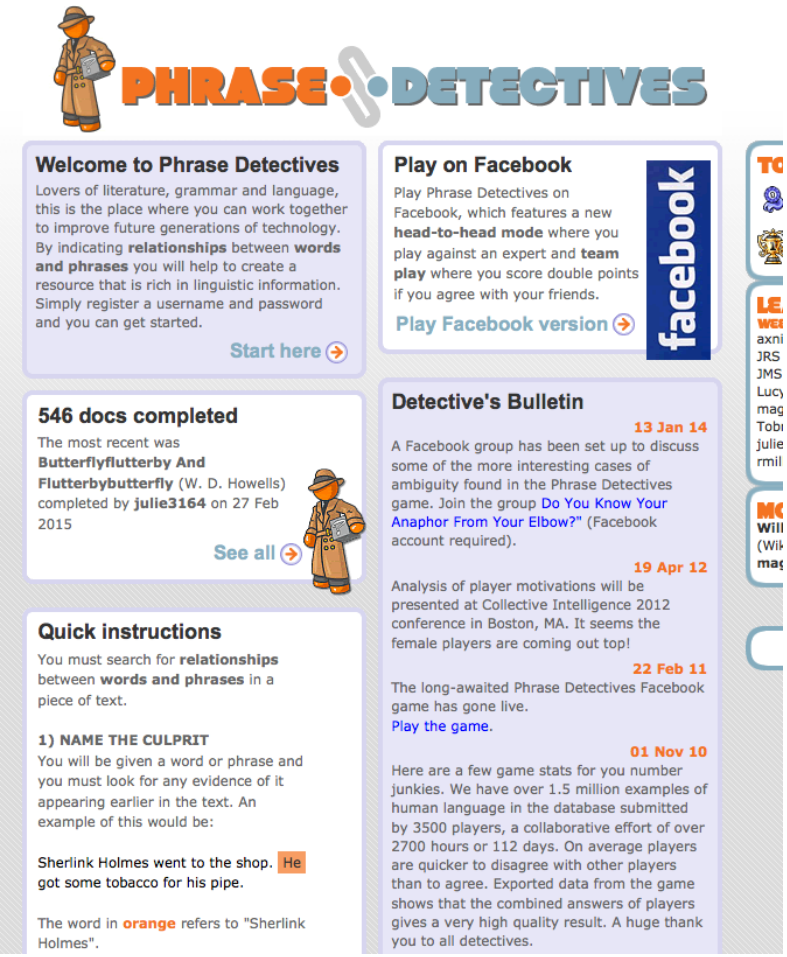
Expert Annotation	1666650	CHF
Crowd-sourcing	2000000	CHF
Actual Total Cost	**> 80000**	**EUR**

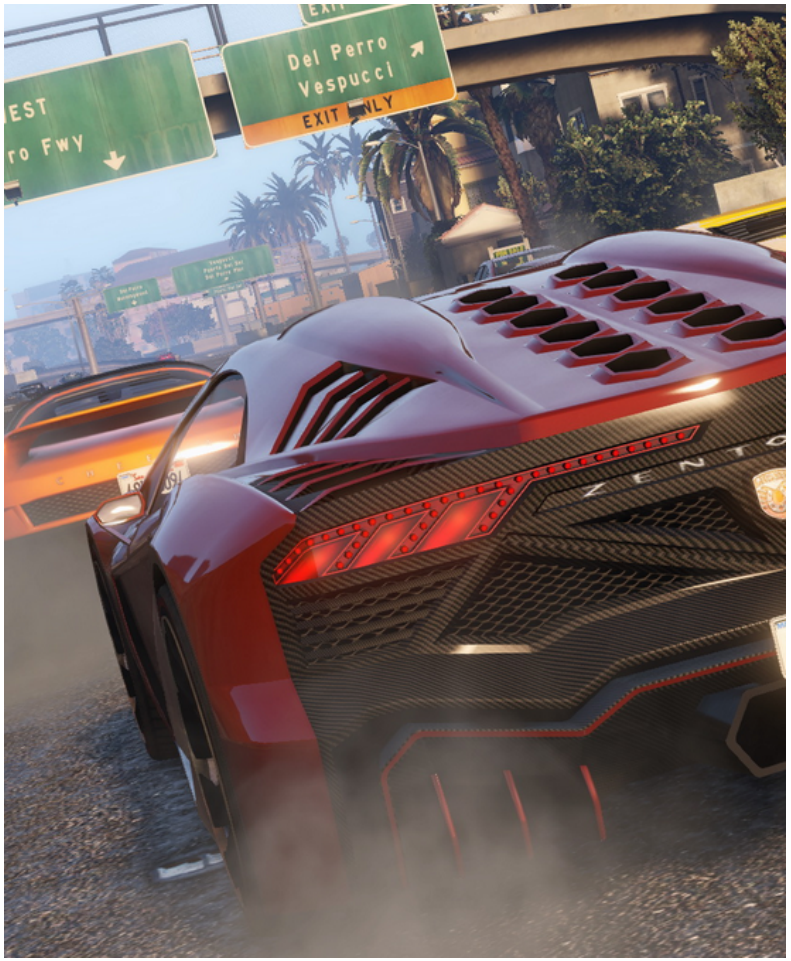
The GWAP approach (at least in the case of JeuxDeMots) is VERY effective

(Mathieu Lafourcade, personal communication)

A *fun*damental problem

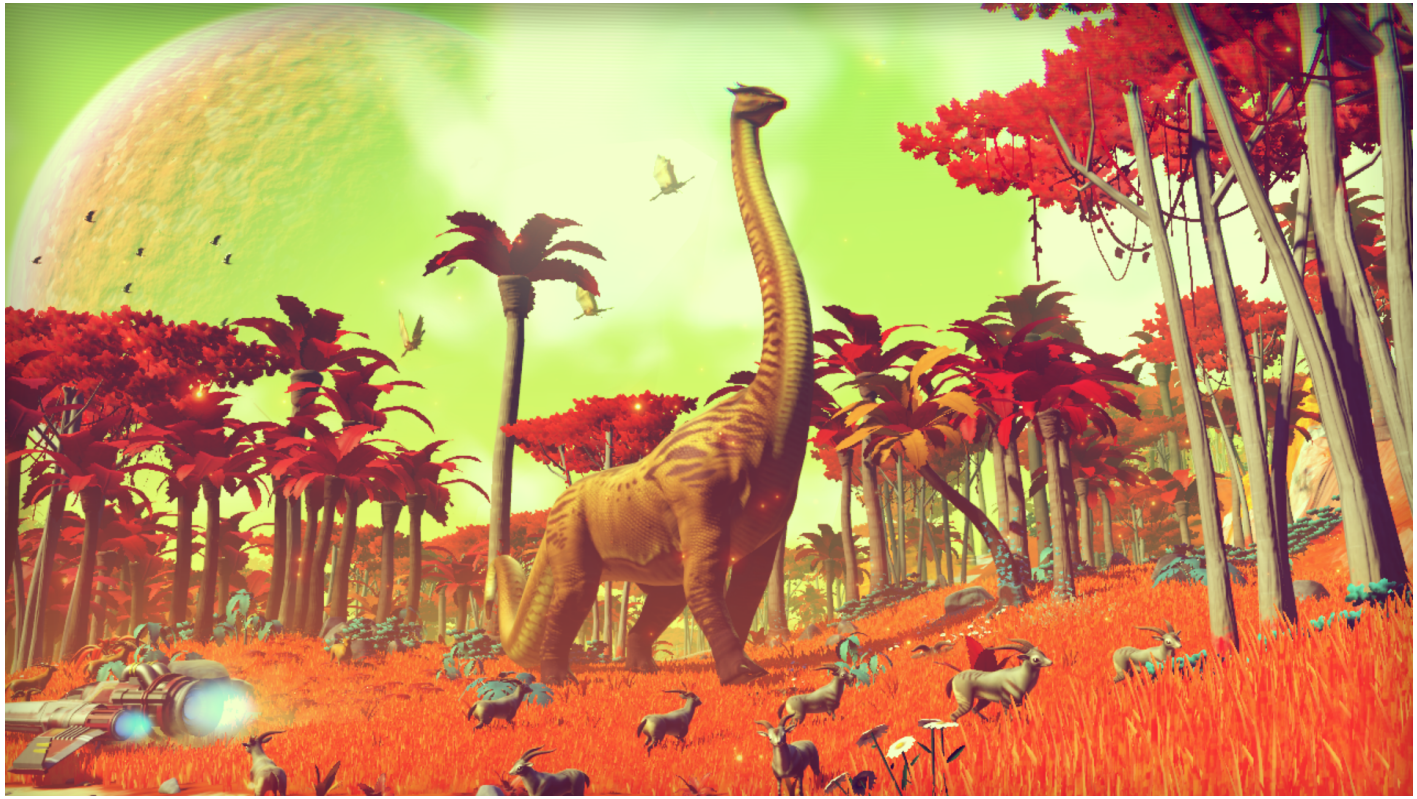
Annotation is not fun.





Vision

- seamlessly integrate annotation tasks in a modern, popular game
- exceptionally interesting: *No Man's Sky*



Bibliography I

Attardi, G. (2010). Phratris – a phrase annotation game. *Unidentifiable Publisher*.

Bry, F., Kneissl, F., Krefeld, T., Lücke, S., and Wieser, C. (2013). A crowdsourcing platform for italian linguistic field research. Research Report PMS-FB-2013-2, Institute for Informatics, University of Munich.

Chamberlain, J., Fort, K., Kruschwitz, U., Lafourcade, M., and Poesio, M. (2013). Using Games to Create Language Resources: Successes and Limitations of the Approach. In Gurevych, I. and Kim, J., editors, *Theory and Applications of Natural Language Processing*, page 42. Springer.

Chklovski, T. (2005). Collecting paraphrase corpora from volunteer contributors. In *Proceedings of the Third International Conference on Knowledge Capture*, K-CAP 2005.

Hladká, B., Mírovský, J., and Schlesinger, P. (2009). *Proceedings of the Third Linguistic Annotation Workshop (LAW III)*, chapter Designing a Language Game for Collecting Coreference Annotation, pages 52–55. Association for Computational Linguistics.

Bibliography II

Jurgens, D. and Navigli, R. (2014). It's all fun and games until someone annotates: Video games with a purpose for linguistic annotation. *Transactions of the Association of Computational Linguistics – Volume 2, Issue 1*, pages 449–464.

Krause, M., Takhtamysheva, A., Wittstock, M., and Malaka, R. (2010). Frontiers of a paradigm: Exploring human computation with digital games. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '10, pages 22–25, New York, NY, USA. ACM.

Kumaran, A., Densmore, M., and Kumar, S. (2014). Online gaming for crowd-sourcing phrase-equivalents. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1238–1247. Dublin City University and Association for Computational Linguistics.

Lafourcade, M. and Joubert, A. (2008). JeuxDeMots : un prototype ludique pour l'émergence de relations entre termes. In *JADT'08 : Journées internationales d'Analyse statistiques des Données Textuelles*, pages 657–666, France.

Bibliography III

Lafourcade, M., Le Brun, N., and Zampa, V. (2014). Colors of people (les couleurs des gens) [in french]. In *Proceedings of TALN 2014 (Volume 2: Short Papers)*, pages 592–597. Association pour le Traitement Automatique des Langues.

Mangeot, M. and Ramisch, C. (2012). *Proceedings of the 3rd Workshop on the People's Web Meets NLP: Collaboratively Constructed Semantic Resources and their Applications to NLP*, chapter A Serious Game for Building a Portuguese Lexical-Semantic Network, pages 10–14. Association for Computational Linguistics.

Pearl, L. and Steyvers, M. (2010). Identifying emotions, intentions, and attitudes in text using a game with a purpose. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, CAAGET '10, pages 71–79, Stroudsburg, PA, USA. Association for Computational Linguistics.

Poesio, M., Chamberlain, J., Kruschwitz, U., Robaldo, L., and Ducceschi, L. (2013). Phrase detectives: Utilizing collective intelligence for internet-scale language resource creation. *ACM Trans. Interact. Intell. Syst.*, 3(1):3:1–3:44.

Bibliography IV

Scharl, A., Sabou, M., Gindl, S., Rafelsberger, W., and Weichselbraun, A. (2012). Leveraging the wisdom of the crowds for the acquisition of multilingual language resources. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*. European Language Resources Association (ELRA).

Siorpaes, K. and Hepp, M. (2008). Ontogame: weaving the semantic web by online games. In *Proceedings of the 5th European semantic web conference on The semantic web: research and applications*, ESWC'08, pages 751–766, Berlin, Heidelberg. Springer-Verlag.

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)*, pages 1294–1304, Baltimore, Maryland. Association for Computational Linguistics.

Venhuizen, J. N., Basile, V., Evang, K., and Bos, J. (2013). *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Short Papers*, chapter Gamification for Word Sense Labeling, pages 397–403. Association for Computational Linguistics.

Bibliography V

Vickrey, D., Bronzan, A., Choi, W., Kumar, A., Turner-Maier, J., Wang, A., and Koller, D. (2008). Online word games for semantic data collection. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 533–542, Honolulu, Hawaii. Association for Computational Linguistics.

von Ahn, L., Kedia, M., and Blum, M. (2006). Verbosity: A game for collecting common-sense facts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 75–78, New York, NY, USA. ACM.