Game Design Evaluation of GWAPs for Collecting Word Associations

Mathieu Lafourcade (1), Nathalie Le Brun (2)

(1) LIRMM, Université de Montpellier
161 rue Ada - 34095 Montpellier Cedex 5 - France
mathieu.lafourcade@lirmm.fr
(2) Imagin@t, France - imaginat@imaginat.name

Abstract

GWAP design might have a tremendous effect on its popularity of course but also on the quality of the data collected. In this paper, a comparison is undertaken between two GWAPs for building term association lists, namely JeuxDeMots and Quicky Goose. After comparing both game designs, the Cohen kappa of associative lists in various configurations is computed in order to assess likeness and differences of the data they provide.

Keywords: lexical network, associative dictionary, game design, comparing association lists

1. Introduction

The aim of the JeuxDeMots[1] project is to design Games With a purpose (GWAPs) to build a large lexical knowledge base (KB). Among the types of relations between terms that structure this network, free association is the one used to infer more precise semantic relations. It should be noted that there has long been a strong interest in associative dictionaries and/or thesauri (like the famous Roget's Thesaurus). What we want to evaluate here is the influence of game design on the quality of the data collected, especially the words that players provide when asked to indicate ideas associated with a target term.

After presenting the lexical network, we compare the main lexical data acquisition game to a new game, Quicky Goose[2] (QG), which proceeds from radically different choices in terms of game design. For a given term, we evaluate the similarity and divergence of the sets of associated terms obtained with each of the game modes, and we try to define the aspects of game design that play a key role.

2. Context: building a large lexical network

In JeuxDeMots (JDM), the main game of the project, players earn and collect words by providing lexical and semantic associations to terms proposed by the system. JDM is a two player GWAP (Game With A Purpose, (Ahn, 2006)) which is both cooperative (a player cannot play "against" another one as at the end of a game all rewards are equally attributed to both players) and competitive (players fight to achieve the best ranking).

Playing games, in order to fill the lexical network, is a kind of indirect crowdsourcing, where people (players) do not negotiate their contribution beforehand. In some cases, direct crowdsourcing (with negotiation between contributors) is desirable. Indeed, some lexical relations might be too complicated to be playable without some linguistic knowledge. That is the case for a telic role, which is the goal/purpose of an object (or action). For instance, a butcher knife has a telic role of cutting meat. It is to be differentiated from the instrument of a predicate, which indicates what can be

done with the object. A butcher knife could be used to stab someone, but this is not its telic role.

2.1. RezoJDM

As mentioned above, the structure of the lexical network (RezoJDM) we are building involves nodes and relations between nodes. Such a structure was initially introduced in the end of 1960s by Collins and Quillian (1969), developed by Sowa and Zachman (1992) and by Fellbaum (1998), used in the small worlds by Gaume et al. (2007), and more recently clarified by Polguère (2014). Every node of the network is composed of a label (which is a term or an expression, or potentially any kind of string), a type (regular term, symbolic information, part of speech, etc.) and a weight, and includes all possible meanings.

The JDM lexical network has a predefined list of around 120 relation types, and around 40 of them are playable in the JDM game. Players cannot define new relation types by themselves. Other games of the JDM project, having a different design, are dedicated to other relations (different from the 40 playable relations of the main game). The JDM relation types fall into several categories:

Lexical relations - synonymy, antonyms, expression, lexical family. These types of relations relate to vocabulary and lexicalization.

Ontological relations - generic (hyperonymy), specific (hyponymy), part of (meronymy), whole of (holonymy), mater/substance, instances (named entities), typical location, characteristics and relevant properties, etc. Such relations concern knowledge about world objects.

Associative relations - free associations, associated feelings, meanings, similar objects, more and less intense (Magn and anti-Magn). These relations are rather about subjective and global knowledge; some of them can be considered as phrasal associations.

Predicative relations - typical agent, typical patient, typical instrument, location where the action takes place, typical manner, typical cause, typical consequence etc. These relations link a verb (or action noun) as starting node to the values of its arguments (in a very broad sense) as ending nodes.

Some relation types are typical of some noun classes or specific domains. For example, for a noun referring to an

---

[1] http://www.jeuxdemots.org/jdm-accueil.php
[2] http://www.jeuxdemots.org/quicky.php

intellectual work (book, novel, movie, piece of art, etc.), the relation "author" is defined. In case of a medical entity, "targets" and "symptoms" relations are defined.

Word senses (or usages) of a given polysemous term T are represented as standard nodes T>glose$_1$, T>glose$_2$, ..., T>glose$_n$ which are linked by refine(ment) relations (of type `r_semantic_raf`) to the term T. Glosses are terms that help the reader to identify the proper meaning of the term T. For example, consider the French term *frégate* (Eng. frigate):

- frégate refine frégate>navire
- frégate>navire refine frégate>navire>ancient
- frégate>navire refine frégate>navire>modern
- frégate refine frégate>oiseau

A frigate can be either a ship or a bird (both English and French show the same ambiguity for this word), and when it is a ship it can be either an ancient ship (with sails) or a modern one (with missiles and such).

As it can be seen from the above example, word refinements are organized as a decision tree, which is more advantageous for lexical disambiguation than a simple list of different meanings.

A particular meaning of a polysemous term is considered a standard term, it can be "played" like any other term. The general polysemous term includes (in principle) the union set of all possible relations for each of the different meanings. In practice, we proceed the other way around, trying to distribute relations from the general term to the proper senses.

Negative Relations - A given relation is weighted, and its weight can be negative or positive. A negative weight is only the result of some contributing process (not possible through the games) where volunteers add information to the lexical network. The interest of negative relations is that they can be at the origin of inhibition processes allowing a semantic analysis system to reject (rather than select) certain meanings during a lexical disambiguation task.

- frégate>navire refine coque
- frégate>oiseau refine<0 coque

If we consider the sentence (in English): *The frigate had her hull breached.* Obviously, the negative relation immediately forbid in this sentence the frigate from being a bird. Thus, negative relations are of primary interest to represent contrastive phenomena among the various meanings of a given term. This later aspect is critical in any approach of lexical embedding. Lexical embedding of words deeply relies on associated information.

2.2. Related Work and Research Questions

Other GWAPs are available to collect word associations. In (Vickrey et al., 2008) three online games are mentioned (Categorilla and Categodzilla and Free Association) that were designed to collect semantic associations in the form of structured data. Users are asked to supply words to fulfill specific categories, for example, "Types of bird", "A thing that cries" etc. The game "Free Association" is based on the popular game Taboo and just asks players to provide words in relation to a target word (stimulus or seed, in the psycholinguistics jargon). There is a taboo list of forbidden words which comes from SemCor and Google unigram data.

The game Verbosity (Ahn et al., 2006) also aims to collect linguistic data and semantic facts. The principle of the game is to propose riddles (a term that a user must make another user guess through the proposal of semantic relations.) In Grác and Nevěřilová (2010), a game similar to Verbosity is presented, but with a strong time constraint of 3 minutes.

In Parasca et al. (2016), an interesting analysis shows how a game can produce data that go beyond automatic extraction based on the distributional hypothesis. The presented game, Word Sheriff, handles word associations as well as more precise semantic relations.

All in all, there are quite no recent GWAPs in NLP to collect word associations. In this context, there are no studies on the effect of game design on data collection. For instance one might ask what effect does the time constraint have on the quality and quantity of data collected? What would be the bias if players could control the proposed target terms ? How to ensure a satisfactory sampling of terms to be proposed without introducing bias?

3. Comparing: two GWAPs for Free Word Associations

We compare the JeuxDeMots and Quicky Goose GWAPs, in order to assess the influence of their respective design on the collection of word associations. Both games concern French language, although they can be adapted to any language.

3.1. JeuxDeMots - A Sophisticated Environment

JeuxDeMots (JDM) was launched late Summer 2007 and since then, more than 1.470.000 games have been played. It should be noted that other games are part of the overall project and contribute to building the lexical network. The user environment is quite sophisticated in order to induce an engagement in players so that they play longer and more often. Although it can be played occasionally, the game is designed to encourage long-term investment, as the goal is to capture, steal, protect and hoard words (which is loosely based on the Pokemon game).

The relevant game design elements of the main game of the JeuxDeMots project are as follows:

- The term to be played with (the target term) is randomly selected;
- the relation (game instruction) is given at the beginning of the game;
- The answers given are compared with those of another player, on the same term with the same instruction. Answers common to both players are added to the lexical network, or strengthen the relation if it already exists.
- Points are computed according to the number of associations common to both players, and the point amount is notified to the user at the end of the game;
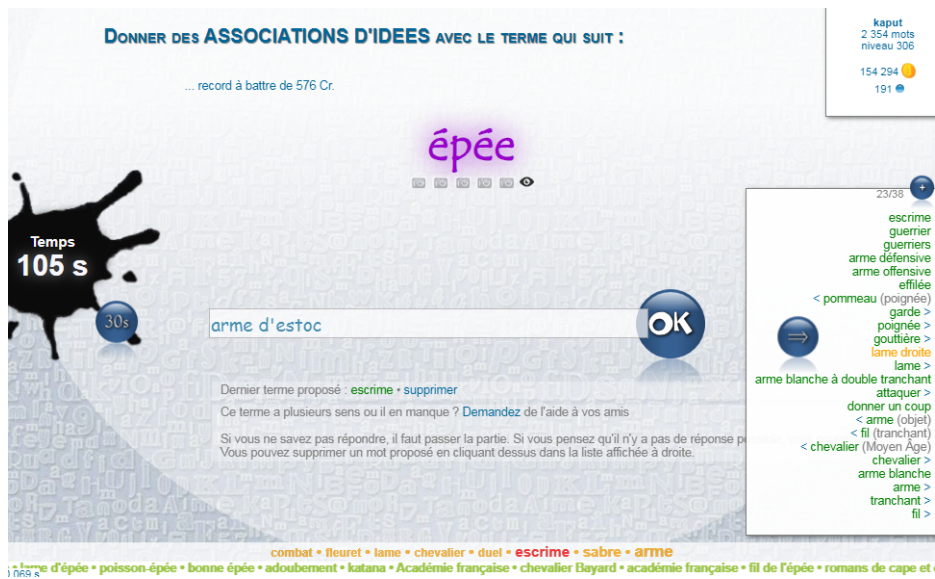
Figure 1: in-going play of JDM. The target word is *épée*. The player has already proposed a number of terms (list on the right).

- a game lasts one minute by default; a timer indicates the remaining time.

Beside elements related to player ranking, JeuxDeMots features *credits*, a virtual money allowing players to have some game control. People can invest directly or indirectly their credits to do the following:

- automatic re-launch of the game (same term and same instruction) which will be confronted with that of another player. This is only allowed below a certain gain threshold.
- buying *competences*, which unlocks the use of more relations
- buy more time while playing;
- investing tokens into his/her last game, so that it will be proposed to more people;
- giving a gift to other players, to encourage them to play with a specific term and relation.

All game features are designed to increase the player's engagement while fostering data quality. The time constraint is an important element in making the game exciting and challenging, although not all players appreciate this constraint: some of them tend to buy a lot of time at the beginning of the game, to reduce stress. And since buying time consumes money, these players are driven to play more, in order to get more virtual money.

3.2. Quicky Goose - A Fast and Simple Direct Approach

Quicky Goose was launched late December 2019, with some word to mouth advertisement on Facebook. Since then, more than 180.000 games have been played, from more than 1000 IP addresses. Registration is not required to play, but someone who is already registered on JDM can play Quicky Goose with his/her account.

The interesting game design elements of Quicky Goose (QG) project are the following:

Figure 2: End of JDM game with display of results. Common words are used for computing points.

Figure 3: End of JDM game with display of results, with alternative results after retry

- The term to be played with (the target term) is randomly selected ;
- The player has the possibility to change relation (instruction) for the current term during the game; but only a subset of all possible relations is available; but a relation that has already been played and then neglected in favor of a new one cannot be chosen again during the current game;
- Answers are compared to the state of the lexical network;

Figure 4: The main screen of the GWAP Quicky Goose. From top to bottom we can see: the instruction, the target term along with current points, the form for entering proposals, and a sum up of already proposed terms. On the left is displayed the *grand total* of points (cumulative points) earned by the player since its first game of QG.

- points are computed on the basis of the presence of each proposed term in the network, either already validated or awaiting validation, or its absence (new proposal); points are added as proposals are made, in real time;
- There is no time limit.

If a term proposed by a player is not already linked to the target term in the network, it is stored in a temporary place. If a certain number of different players (empirically set to 5) propose the same association, then it is added to the lexical network. If they are not registered, players are identified by their IP address. Thus, the same proposal cannot be given several times by the same player. The creation weight is set to 5 and each subsequent association increases it by 1. As such, the weight strictly translates the number of times two terms have been associated. In JDM by contrast, a new association enters the lexical network if at least two players have made this association and met by chance during a game (player A is playing on the recorded game of player B who already made this association). As the other player is not known during the game, collusion is not possible.

Displaying points earned for each proposed word during the course of the game (and not at the end) makes the game highly addictive for some people. Players tend to beat their own record (proposing more terms, earning more points). Although QG is considered a purely casual game, the average playing time tends to exceed that observed on JDM, because as people are not subject to the time constraint, they continue to make proposals until they have no more ideas.

Selection of target words - QG uses the JDM lexical network (which is in open access) to select a target word to propose to the player. By default, QG selects common words (this is a possible attribute of network terms in Rezo-
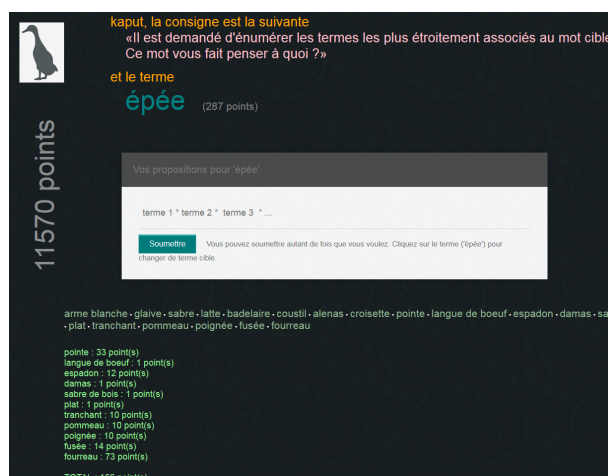

Figure 5: The proposed terms are displayed below with points earned for each one. Once a term is proposed, is cannot be proposed again.

JDM) but can alternatively propose target terms according to a theme, freely chosen by the player: then only words linked to the theme in RezoJDM are proposed. It should be stressed that a purely random word selection from a pre-defined list is not suitable, and would tend to make player either bored or desperate because very unusual or improbable target terms are then proposed to him/her. So it is preferable, when possible, to use a preexisting knowledge base.

Having plenty of time - with QG, players have all the time they want to propose word associations. For difficult relations, such as "telic role", this aspect of the game is welcome since the player can think about his/her answers, their quality, relevance and diversity increases.
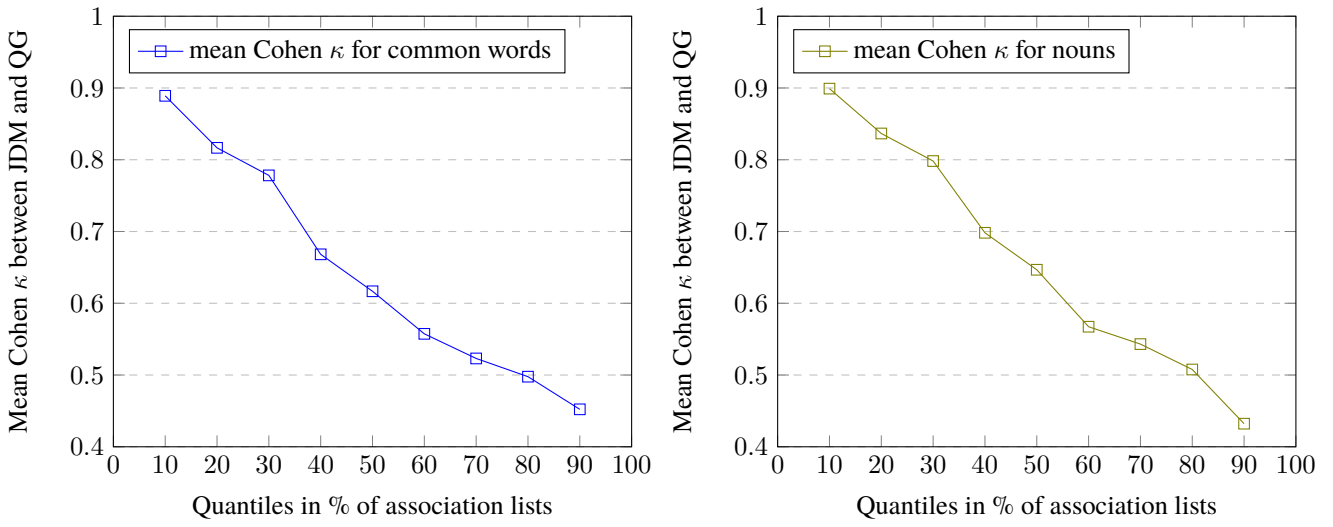
29

Figure 6: (left) mean Cohen κ between associative list of JDM and QG for 2800 common words (Nouns, Verbs, and Adjectives) for free associations; (right) mean Cohen κ between associative list of JDM and QG for 3200 Nouns for free associations ;
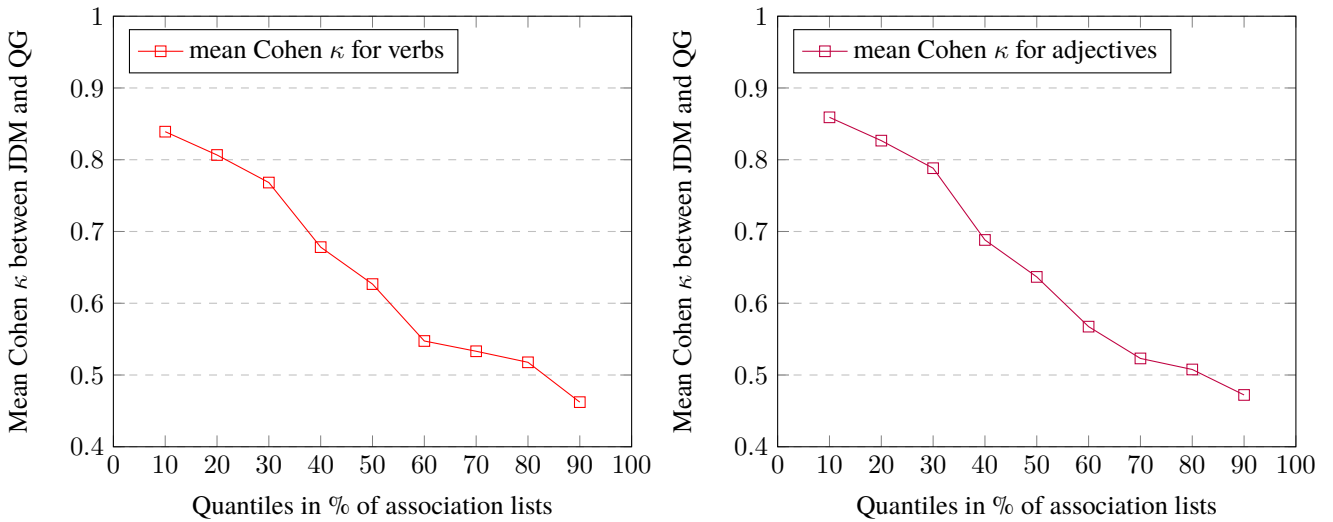

Figure 7: (left) mean Cohen κ between associative list of JDM and QG for 890 verbs for free associations ; (right) mean Cohen κ between associative list of JDM and QG for 1450 adjectives for free associations.

4. Evaluating: Agreement Between Associative Lists and Effects of Game Designs

We evaluated the associations made by the players, both in JDM and in QG. What are the similarities and differences between the two game modes in terms of the data collected? What is the value of data collected using one of the two game modes, that are not collected using the other ? Which game features induce these differences?

		B	
		Yes	No
A	Yes	a	b
	No	c	d

Figure 8: Situations for the Cohen κ (from Wikipedia). In our evaluation the d situation is not possible and thus is always equal to 0. At least one list contains the tested term.

4.1. Quantitative assessment

The methodology for evaluating agreement is as follows: for a given target word T, we took both association lists produced with JDM and with QG respectively. We can compute an agreement (Cohen κ) for quantiles of the list. We adopted an approach with 10 quantiles. The first quantile corresponds to the first (most activated) 10% terms of the association list. The second quantile, are terms ranked between 10% up to 20% (not included), and so on. Choosing the number of quantiles was a difficult question. In many studies quartiles are used (four quantiles of 25%), but in our case it was to be a bit coarse. One of the motives for a finer quantification (using 10 quantiles of 10% rather than quartiles) is that the data behave according a power law and not an average distribution, and as such the variations of the distribution are much stronger at the beginning (first quantiles).

The value of the agreement of the Cohen κ is a global measure of the answers to questions like "is the word A present in the n% quantile of associations for relation t for term B ?" For example, "is *mouse* present in the 10% quantile of associations for relation r_associated (free associations) for *cat* ?" Formally we ask each association list (from JDM and QG) and their answer is either "yes" or "no". Note that case with both answering "no" is not possible as we compare both lists, hence a given term is necessarily at least in one of the lists.

The domain of the agreement is every time the union of terms of both lists, and not all the possible terms existing in JDM. If we had proceeded this way, the agreement would always have been meaninglessly close to 1, because there is an overwhelming number of terms no whatsoever related to each other. In our evaluation the d case in figure 8 is always equal to 0.

This assessment evaluates only the rank of the associated words, not their actual weight. The weight is only significant when comparing terms within the same list. Hence relying on weights to compare two lists produced through different means would be meaningless. Also, knowing the exact rank of a term in one list in order to compare it to its rank in the other list is not really meaningful. What is required for evaluation is just whether the two terms are in the same part of the distribution curve.

In figure 6 it appears that the agreement between JDM and QG association lists is very high in the first quantiles. That is to say, that the strongest associations are very similar (if not identical). As intuitively expected, the Cohen kappa decreases as the rank of the associations increases (their strength is decreasing). We can see that the global highest agreement concerns nouns, adjectives, and finally verbs. Indeed, finding association to verbs is felt to be not quite so easy, as people often focus on synonyms and potential patients (eat an apple). But more often than not, the number of possible answers is high, hence a lower agreement than for nouns. For adjectives, beside synonyms and antonyms, the more recurrent associations are the typical targets (red apple, red car, red skirt, etc.) even though the "complete" association list is both very large and illusory to achieve.
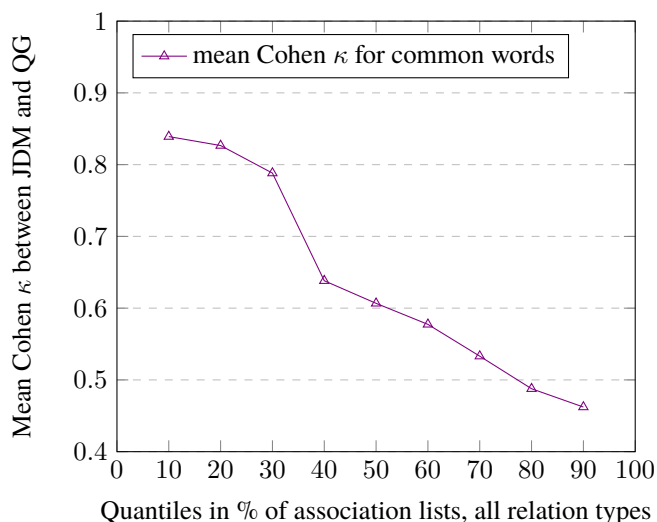
Figure 9: Mean Cohen κ between associative list of JDM and QG for 2800 common words (Nouns, Verbs, and Adjectives) for any relation type, except for free associations.

The agreements for other relations than associated ideas (figure 9) are not fundamentally different than the ones for associated ideas. What we can notice the Cohen κ agreement is very strong up to the first 30%, then drops sharply to around 0.63. The main reason is that in general for precise semantic relations there are fewer obvious possible answers than for free associations. Hence the agreement between JDM and QG is lower because due to the time constraint people usually don't have time to propose a high number of relevant answers. Some players "confessed" that since they have unlimited time in QG, they consult online encyclopedias and dictionaries or even hard books. Given the purpose of the project (building a lexical semantic resource) we can only approve of such behavior. Some might mock such players as "information extractors from external existing resources" to be opposed to players that are "information extractors from internal existing resources" (namely their brain).

4.2. Qualitative assessment

For the qualitative assessment, we tried to evaluate two aspects: a) number of false or dubious terms in associations, b) the quality of terms in disagreement in association lists (i.e those belonging to only one of the two lists).

Errors - We looked for terms that are not common to both lists and semi-automatically evaluated whether they should be considered errors. For over 7000 associations lists (over 3500 for JDM and 3500 for QG) corresponding to about 500.000 terms, we found around 1% terms (around 5000 "rogue" terms) than were not in both lists. We first made a random manual evaluation of around 500 terms (10%) and did not find any wrong, and only 13 that might be considered as far fetched. For example, *chat* (cat) and *Alice* are associated (probably because of their link with the *Cheshire Cat*) has been considered as far fetched.

In a more systematic way and by exploiting the RezoJDM, we tried to (automatically) assess if a rogue term A could be indirectly linked to the target term B through an intermediate term C. For example, *pavé* is linked to *main* in the QG association list but not in JDM (not in rezoJDM, hence). But in rezoJDM, we can have *pavé* (A) linked to *lancer* (C) linked to *main* (C). Thus, we consider the association between *pavé* and *main* to be correct (which is the case). With this method, only 25 rogue associations were found not linkable through an intermediate nodes. By manually checking those associations, 17 were considered as correct and 8 as far fetched. None were considered as clearly false.

Undoubtedly, there are (of will be) errors in associations, nevertheless they seem to be quite uncommon and hard to spot automatically, and even human judgment might be difficult in some cases. The JDM filtering (2 players encountering) and the QG filtering (5 independent contributions) appears to behave as expected reducing the amount of errors entering the KB (rezoJDM). During the development of these games, theses respective number (2 and 5) of different users for confirming an association have been determined empirically, trying to have a good balance between recall and precision.

Terms in disagreement - It is difficult to compare two lists that were constructed on different time scales (13 years for JDM, and 2 months for QG). However, it is possible to normalize the heaviest list (with the highest strongest score) to the weakest by linearly reducing weights to the weakest. The reduction is linear as all weights are divided by the maximum weight.

After normalization, comparison shows that QG association lists are richer than those of JDM, but tend to be much "flatter" (smaller relative difference between term weights). In contrast, JDM association lists have higher weight variations (more contrast), even if globally term ranks are the same for the first quantiles.

Again, in QG people tend to be very creative, proposing quite often terms that do not yet exist in rezoJDM, but in most cases are very relevant.

4.3. Impact of the Game Design Choices

We did a very small survey by asking the identified players about the features of QG and/or the features of JDM. Their
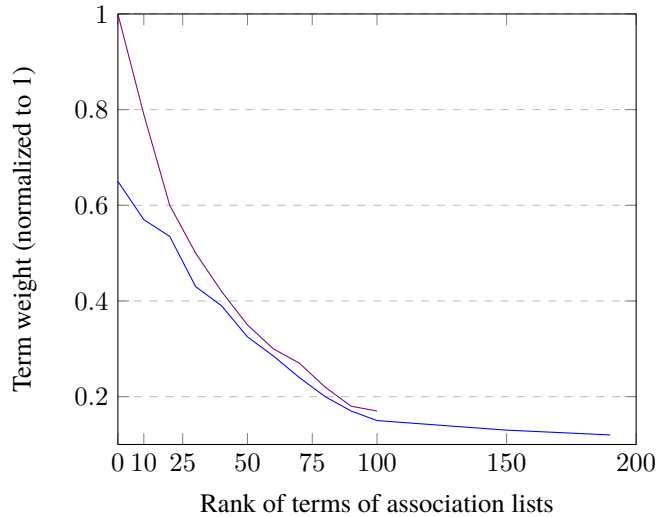

Figure 10: Illustration of typical association lists (JDM purple, QG blue), weights are normalized to 1. The QG plot is flatter but longer than the JDM plot.

answers are consistent with our evaluation of association lists.

The time constraint in JeuxDeMots has a very strong impact on the data collected. It makes the game exciting for many people, but some people find it stressful. Without a time constraint, players produce more associations, so that lead to a quite longer tail of associations. However, the collected data are a bit less spontaneous. It does however not have much impact on the strongest associations.

The immediate display of points gained during the game (like in QG) encourages longer play time and proposing more associations. In JDM, getting the result only at the end of the game makes it like a kind of bet, leads to either excitement or disappointment. In the overall, players produce more associations when the points are distilled during the process. True players play JDM, and most don't care to contribute but instead are obsessed with ranking, and other rewards. People minded to contribute plays QG, even if they think they are players.

Being able to change the instruction during the game seems to be appreciated by players (whether they already played JDM or not). The (little) constraint of not allowing to propose again an already proposed term, even with another instruction is controversial. Some people think this is an unnecessary restriction, others see that as a game challenge. After playing a little, some players keep some associations for the most appropriate instruction. The set and order of instructions is defined along with the nature of the term (noun, verb, adjective, adverb, ...) and with some experience the player does anticipate the "most appropriate place" where the candidate association belongs. This aspect makes the game challenging and just a direct "put all of them here" kind of activity. Furthermore, for data quality sake, this feature induces players to properly distributed associated terms among semantic relations.

The number of players (after scaling) of QG is much higher than JDM (> 1000 in two months for QC, and > 10000 in 13 years for JDM), even if the novelty effect is taken into

account. The simplicity of QG tend to favor the number of players, but the turnover is higher than in JeuxDeMots (6 days versus 24 days). The distribution of players according to their number of games done, follows in both case a power law (few people playing a lot, and most people playing occasionally).

5. Conclusion

In this paper we have presented a comparison between two GWAPs for building term association lists, namely JeuxDe-Mots and Quicky Goose. After comparing game design in both games, we computed the Cohen kappa of associative lists in various configurations in order to assess major differences in obtained data.

It appears that game with a time constraint is more exciting for many but tend to produced less flourishing associative lists than games without this time constraint. This tendency is only noticeable for the long tail of associative lists, that is for less activated associated terms. Some people just prefer having as much time as they want, and collecting points and rewards this way.

We do not know yet the percentage of people who are playing both games, or whether the gamer population is strictly separated. Anyway, it seems quite clear that proposing several games with different designs in the context of the same project is a good (if not cheap) strategy for building a valuable linguistic resource.

6. Acknowledgments

7. Bibliographical References

Ahn, L. V., Kedia, M., and Blum, M. (2006). Verbosity: a game for collecting common-sense facts. In In Proceedings of ACM CHI 2006 Conference on Human Factors in Computing Systems, volume 1 of Games, pages 75–78. ACM Press.

Ahn, L. v. (2006). Games with a purpose. *Computer*, 39(6):92–94, June.

Collins, A. M. and Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8(2):240 – 247.

Fellbaum, C. (1998). WordNet: An Electronic Lexical Database. Bradford Books.

Gaume, B., Duvignau, K., and Vanhove, M. (2007). Semantic associations and confluences in paradigmatic networks. In From polysemy to semantic change - towards a typology of lexical semantic associations, Toulouse, France. John Benjamins Publishing Company.

Gaume, B., Ho-Dac, L. M., Tanguy, L., Fabre, C., Pierrejean, B., Hathout, N., Farinas, J., Pinquier, J., Danet, L., Péran, P., et al. (2019). Toward a computational multidimensional lexical similarity measure for modeling word

association tasks in psycholinguistics. In Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, pages 71–76.

Grác, M. and Nevěřilová, Z. (2010). Acquiring nlp data by means of games. pages 109–114, 01.

Lafourcade, M., Joubert, A., and Brun, N. (2015). Games with a Purpose (GWAPS). Focus Series in Cognitive Science and Knowledge Management. Wiley.

Lafourcade, M., Mery, B., Mirzapour, M., Moot, R., and Retoré, C. (2018). Collecting Weighted Coercions from Crowd-Sourced Lexical Data for Compositional Semantic Analysis. In isAI: International Symposium on Artificial Intelligence, volume LNCS of *New Frontiers in Artificial Intelligence*, pages 214–230, Tokyo, Japan, June.

Lafourcade, M. (2007). Making people play for Lexical Acquisition with the JeuxDeMots prototype. In SNLP'07: 7th International Symposium on Natural Language Processing, page 7, Pattaya, Chonburi, Thailand, December.

Machida, Y., Kawahara, D., Kurohashi, S., and Sassano, M. (2016). Design of word association games using dialog systems for acquisition of word association knowledge. In Proceedings of the 5th Workshop on Automated Knowledge Base Construction, pages 86–91.

Mitchell, T., Cohen, W., Hruschka, E., Talukdar, P., Yang, B., Betteridge, J., Carlson, A., Dalvi, B., Gardner, M., Kisiel, B., Krishnamurthy, J., Lao, N., Mazaitis, K., Mohamed, T., Nakashole, N., Platanios, E., Ritter, A., Samadi, M., Settles, B., Wang, R., Wijaya, D., Gupta, A., Chen, X., Saparov, A., Greaves, M., and Welling, J. (2018). Never-ending learning. *Commun. ACM*, 61(5):103–115, April.

Parasca, I.-E., Rauter, A. L., Roper, J., Rusinov, A., Bouchard, G., Riedel, S., and Stenetorp, P. (2016). Defining words with words: Beyond the distributional hypothesis. In Proceedings of the 1st Workshop on Evaluating Vector-Space Representations for NLP, pages 122–126, Berlin, Germany, August. Association for Computational Linguistics.

Polguère, A. (2014). From writing dictionaries to weaving lexical networks. *International Journal of Lexicography*, 27(4):396—418.

Singh, P., Lin, T., Mueller, E. T., Lim, G., Perkins, T., and Li Zhu, W. (2002). Open mind common sense: Knowledge acquisition from the general public. In Robert Meersman et al., editors, On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE, pages 1223–1237, Berlin, Heidelberg. Springer Berlin Heidelberg.

Sowa, J. F. and Zachman, J. A. (1992). Extending and formalizing the framework for information systems architecture. *IBM Systems Journal*, 31(3):590–616.

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 Conference on Empirical Methods in Natural Language Processing, EMNLP '08, page 533–542, USA. Association for Computational Linguistics.