HAL
open science

Ambiguss, a game for building a Sense Annotated Corpus for French

Mathieu Lafourcade, Nathalie Le Brun

To cite this version:

Mathieu Lafourcade, Nathalie Le Brun. Ambiguss, a game for building a Sense Annotated Corpus for French. IWCS 2017 - 12th International Conference on Computational Semantics, Sep 2017, Montpellier, France. lirmm-01763434

HAL Id: lirmm-01763434

https://hal-lirmm.ccsd.cnrs.fr/lirmm-01763434v1

Submitted on 11 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Ambiguss, a game for building a Sense Annotated Corpus for French

Mathieu Lafourcade
LIRMM, 161, rue ADA 34392
Montpellier Cedex 5 - France
lafourcade@lirmm.fr

Nathalie Le Brun
Imagin@t
France
imaginat@imaginat.fr

Summary : This paper presents Ambiguss, a Game With A Purpose designed both to collect ambiguous sentences and build a Sense Annotated Corpus. It also generates a lexicon of polysemous words associated with the glosses that illustrate the different meanings. Early evaluations indicate that the approach is relevant and efficient.

1 Introduction

Evaluating a WSD task is a challenge at least as difficult as developing the task itself. Manually constructing a corpus of ambiguous sentences is a difficult and tedious task. Such a corpus is even more complex to produce if in each sentence, each ambiguous word is associated with its correct meaning. In addition to finding/imagining the sentence, it is necessary to have a word sense lexicon to associate the correct meaning with each ambiguous word, and such a resource may not be available in any given language. Secondly, referring to a given meaning might be tricky. This can be done by associating a meaning number, but in such a case, the lexicon must be provided along with the corpus. Another method would be to represent the correct meaning of the ambiguous word by a gloss, i.e. a word or group of words that intuitively refers to its correct meaning. A classic example could be *bank > river* and *bank > money*; the glosses *river* and *money* refer to two possible meanings of the word *bank*. Anyway, to define the glosses, i.e. to choose the word that best illustrates the meaning of a word is another pitfall: Indeed, it seems that there is not always a strong agreement between people for such a task.

Collect lexical data by crowdsourcing through games is a new trend since few years. The question that arises is to properly identify what kind of data can be collected in the domain of Natural Language Processing and Semantic computing. The postulate is that the semantic information understandable by the native speakers can be collected in this way. Furthermore, the type of semantically ambiguous sentence in which we are interested here, refers mostly to common sense knowledge (to be tackled with knowledge bases in the spirit of (Lenat, 1995)). This type of knowledge is intuitive and easy to infer for speakers, but definitively difficult to model to develop an Artificial Intelligence that would perform properly WSD.

This paper presents the design of a Gwap (Game With A Purpose) called *Ambiguss*, whose goal is twofold. On one hand, this game allows collecting ambiguous sentences created by players. On the other hand, these sentences are proposed to other players who must select from a list of glosses the one that corresponds to the correct meaning of each ambiguous word of the sentence. First, we show that the concept of Gwap is adapted to the production of Sense Annotated Corpus. Then we describe a novel Gwap designed to build a Sense Annotated Corpus. We conclude with a shallow evaluation of the first data collected.

2 From GWAPs to Sense Annotated Corpora

How could we relate and apply the idea of Gwap to the production of a Sense Annotated Corpus?

With the advent of Gwaps and more generally crowdsourcing approaches, one observe a fundamental shift in language resources building that is going on now (see (Chamberlain *et al.*, 2013) for more details). If the use of crowdsourcing microworking platforms like Amazon Mechanical Turk can be questionable, both in terms of ethics and quality (Fort *et al.*, 2011), Gwaps represent a paradigm shift

for the acquisition of lexical resources, which passes from hand-made by specialists to crowdsourced through entertainment by native speakers. However, creating an efficient GWAP involves principles and developments that are not that straightforward.

The concept of Gwap was first proposed by (von Ahn and Dabbish, 2004 and 2006) and is based on the idea that beyond the simple entertainment of players, a game can have a purpose, such as data acquisition or problem solving. Since then, the first principles defined by (von Ahn *et al.*, 2008) have given rise to Gwaps in various fields (see Lafourcade, *et al.*, 2015 for a survey). In (Chamberlain *et al.*, 2009), the Phrase Detectives game is proposed to solve coreferences in texts, which is conceptually quite close to construct a corpus of sense annotated sentences. Wordrobe (Venhuizen *et al.*, 2013) is a game for labeling word senses, but senses come from a predefined set (the Groningen Meaning Bank).

Besides being funny and addictive, the mandatory characteristics for a Gwap to be successful could be found by asking simple questions: what can we ask to players? How could we ask these things to players? The task of selecting the proper meanings of words contained in sentences is clearly within the reach of native speakers. But one may wonder if this is a fun enough task to give rise to a game, and entertaining enough so that players are willing to do it?

According to (Habert et al., 1998) a corpus is a collection of language related data which are selected and organized according to linguistic criteria so as to be used as samples (for evaluation, for instance). Amongst those corpora, there are the British National Corpus (Burnard, 1998) (100 million words) and the American National Corpus (Ide & Macleod, 2001) (20 million words). In a Sense Annotated Corpus each word is related to a word meaning, usually by means of a sense number linked to a specific database (see Vial *et al.*, 2017). For example, the corpus of the task number 7 of SemEval 2007 is annotated with the identifiers of Princeton WordNet 2.1 (Fellbaum and Miller, 1998). The English corpus for the task 13 of SemEval 2015 2015 (Moro & Navigli, 2015) is annotated with the identifier of version 3.0 of BabelNet. Building such corpus might be based on various motivations, the most obvious of them being to provide a basis for automatically evaluating WSD systems.

For the French language, there is no such corpus freely available. Furthermore, there is a strong need for a corpus that would not be directly dependent on a given database. Hence the interest of indicating the meaning of words by glosses rather than identifiers related to specific meanings of a specific resource. Evaluating the feasibility of developing a new GWAP designed to construct such a corpus sounds interesting. The main issues lie in the design of the game which must be addictive and efficient, and secondarily in the evaluation of the collected data.

3 Ambiguss, a game to collect Semantically Ambiguous Sentences

Ambiguss (https://ambiguss.calyxe.fr/) is a game to solve lexical ambiguities in sentences, but also to collect such semantically ambiguous sentences (SemAS). Although this might be the case with some collected sentences, Ambiguss is not designed to collect syntactically ambiguous sentences. Everything is designed so that players do only need common language knowledge. The game is aimed at native speakers, who are not expert in linguistics or grammar.

3.1 Playing ambiguous sentences

Once logged, the player is proposed a random sentence in which at least one word is highlighted. The purpose of the game is to select the proper meaning of each highlighted word, through an user interface menu that presents a list of several possible meanings. When done, the player has to click on the *Valider* (Eng. Validate) button, and then he gets the result (Figure 2). There is no time constraint in the game.

In Figure 1, an example of game is given. The words *rat* and *radis* are highlighted: this means they are polysemous and the player is invited to disambiguate them by choosing the correct meaning according to the sentence, as seen in the figure for the word *rat*, which has two possible meanings.

Figure 1: An Ambiguss play with the sentence *Ce rat ne prête pas un radis*.

Figure 2: A screenshot of the game result. The player earns the same number of points and credits, here 100.

Players earn two types of rewards, *points*, only for fame, and *credits* that can be used for typical game actions, such as creating oneself an ambiguous phrase or adding a gloss: indeed, during the game, a player may add his own gloss, if he feels that none of the proposed ones is appropriate. The amount of points earned is related to the distribution of other player answers. (Of course, the main hypothesis, which is verified experimentally, is that *globally*, players provide the adequate answers).

3.2 Creating ambiguous sentences

The opportunity to create an ambiguous sentence is both a reward for good players and an incentive to play well, while being a way of spending the credits earned by playing. To do this, the player has to select the *"Créer"* menu and key in a sentence. He/she can then select one or more word(s) in this sentence and declare them as ambiguous. If an ambiguous word does not yet have glosses, he/she adds some to it to establish a list of possible meanings, and indicates what the correct gloss is for the term in the context of his/her sentence. These ambiguous words are those that will be highlighted and that another player will have to disambiguate in a game.

Although the user interface was deemed intuitive by the players, for now, only a small fraction of them (15%) has created sentences. It would seem that the fear of being judged through the quality of his own sentences might be deterrent to some players.

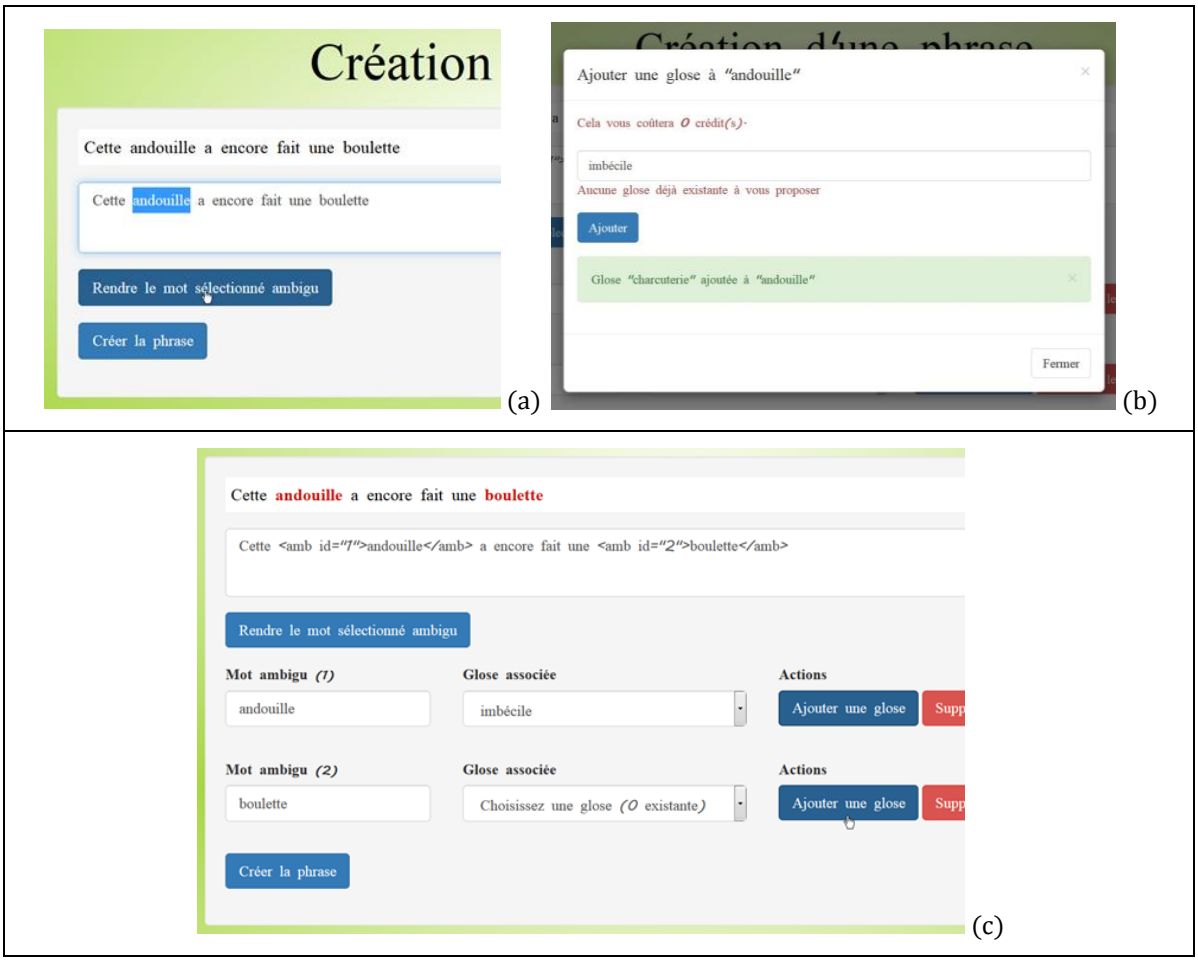

Figure 3: Screenshots of the different steps (a), (b), (c) of creating an ambiguous sentence. (a): Enter the sentence and select the ambiguous words. (b): add glosses for each ambiguous word. (c): Choose the correct gloss for each ambiguous word.

Why creating sentences? For greed and fame of course!

Each time a player earns *points* and *credits*, the creator of the sentence recovers 10% of the total. Hence, from a strict gaming point of view, players have a definite interest in creating sentences with many interesting ambiguities. When he/she plays, a player can pass over a sentence, if he/she finds it boring or nonsensical. Moreover, when playing, people can *like* a sentence (and share it on Facebook and Twitter). Having a high number of likes is an incentive for many players. Hence, they trend to produce interesting and funny sentences. Some of those sentences are "undecidable" in the absence of context, for example:

Je suis une fille.	*L'avocat est véreux*	*La petite brise la glace*	*Je loue un appartement*

In such cases, the interest of the game is to make emerge the preferred meaning, the one chosen by the majority of the players. But in some cases, a quite even number of votes for glosses is also possible.

The game proposes a multi-criteria ranking that orders players according to various parameters (number of points, number of sentences created, and number of *likes* obtained). Another ranking concerns the sentences and orders them according to the number of likes or the gains they have given to their creator. All this is designed to flatter and retain the player and thus intensify his contribution.

3.3 About the data collected

Ambiguss is still a prototype, but people have been asked from the beginning to serve as testers. By word of mouth, we reach after few weeks a total number of 64 players. Most of them have just played the disambiguation game, few daring to embark on the creation of sentences. A manual evaluation by

expert in NLP concluded that 96% of the sentences are valuable for WSD evaluation as they are mostly good examples of semantic ambiguity related to common sense.

Number of sentences	354
Number of ambiguous words	524
Number of glosses	1556

Table 1: The amount of data collected with around 60 players in a few weeks.

Are the sentence properly disambiguated? Are the collected glosses correct?

So far, 100% of the sentences are disambiguated properly. More precisely, the gloss predominantly selected is the one that corresponds to the correct meaning in 100% of the cases. If we go into details, players achieve collectively a precision of 0.985. This means that it is very rare for a player to select a wrong meaning. Recall is not applicable in this context. Wrong glosses for a given term are never selected, but some glosses (4%) are not very representative of the meaning to be illustrated.

Reporting incorrect data

Players have access to an user interface button to indicate an incorrect item, whether it's a spelling mistake in a sentence or a gloss, a bad gloss, an absurd phrase, an ethical violation, and so on. The reporting mechanism allows self healing of the data and is very fast in practice (wrong data do not stay very long uncorrected).

3.4 Exporting the data

The data collected with Ambiguss are freely available to anyone. Two sets of data are available: the sentences, and the list of ambiguous terms with their glosses. The export format is JASON. For example, exporting the sentence *C'était un vol agréable malgré la distance* (Eng. It *was a pleasant flight/theft despite the distance.*) produces the following output :

{"phrase":"C'était un <amb id=\"1\">vol<\/amb> agréable malgré la distance.","reponse":[{"motAmbigu":"vol","ordre":1,"nbRep":30,"gloses":[{"valeur":"crime","nbRep":"1"},{"valeur":"voyage","nbRep":"29"}]}]}

The sentence is given with inside annotations for indicating the ambiguous words. The field *"réponses"* (Eng. *Answers*) gives for each ambiguous word the distribution of the glosses proposed by the players. In the above example, 29 players answered *"voyage"* for the word *"vol"* and only 1 player proposed *"crime"*. Of course, this particular sentence can hold at least two interpretations, but only one sounds reasonably meaningful. Concerning the lexicon of terms and glosses, the export is also under a JASON form:

{"motAmbigu":"vol","gloses":["voyage","crime"]}
{"motAmbigu":"souris","gloses":["animal","objet","gigot","périphérique informatique","nana"]}

For the word *souris*, 5 glosses are proposed. It is most than probable that in many cases more glosses are proposed than the actual number of meanings. That is to say, for some meaning, several glosses are present. So, there would remain a task not foreseen by the game: to distribute the glosses for each polysemous word in equivalence classes. For example, in the case of *souris*, both glosses « *périphérique informatique* » and « *objet* » might probably refer to the same meaning.

4 Conclusion

We presented a game, called Ambiguss, designed for collecting ambiguous sentences along with player choices for the proper meanings. Another output of the game is that it builds a lexicon of polysemous words with their associated glosses. Although still in its testing phase, the game was judged *highly addictive* by many testers.

We have already collected several hundred semantically ambiguous sentences (for French). This corpus can be freely used as a reference for easily evaluating an automatic WSD process (or any relevant tasks). Word meanings are represented as glosses, which avoids the use of a dedicated semantic lexicon. The list of polysemous words and glosses is also freely accessible.

Acknowledgements

The Ambiguss game would not have existed without the dedication of the members of the TER M1 (Travail d'Etude et de Recherche, Master 1 en Informatique) student group IMNA, namely Isna, Melissa, Nicolas, and Alexandre. A big up for them.

References

VON AHN L., AND DABBISH L. (2004). *Labelling Images with a Computer Game.* ACM Conference on Human Factors in Computing Systems (CHI). pp. 319-326, New York, NY, USA, 2004. ACM Press.

VON AHN, L., KEDIA, M., AND BLUM, M. (2006). *Verbosity: a game for collecting common-sense facts.* In Proceedings of ACM CHI 2006 Conference on Human Factors in Computing Systems, volume 1 of Games, pages 75–78, 2006.

VON AHN L., AND DABBISH L. (2008*). Designing Games With a Purpose.* Communication of the ACM, 51(8):58–67 August 2008.

BASILE V., BOS J., EVANG K., AND VENHUIZEN H. (2012). Developing a large semantically annotated corpus. *LREC 2012, Eighth International Conference on Language Resources and Evaluation*, May 2012, Istanbul, Turkey.

BURNARD L. (1998). *The British National Corpus*, The digital word, MIT Press Cambridge, MA,1998.

CHAMBERLAIN J, POESIO M, AND KRUSCHWITZ U. (2009). *A demonstration of human computation using the Phrase Detectives annotation game*. In HCOMP '09: Proceedings of the ACM SIGKDD Workshop on Human Computation pp. 23-24.

CHKLOVSKI T., AND GIL Y. (2005). *Improving the design of intelligent acquisition interfaces for collecting world knowledge from web contributors.* In Proceedings of K-CAP '05, pages 35–42.

FELLBAUM C. AND MILLER, G. editors. (1998). *Word-Net*. The MIT Press, 1998.

HABERT B., FABRE C. AND ISSAC F. (1998). *DE L'ECRIT AU NUMERIQUE. Constituer, normaliser et exploiter les corpus électroniques.* Number ISBN : 2-225-82953-5. Elsevier Masson.

FORT, K., ADDA, G., AND BRETONNEL COHEN, K. (2011). *Amazon Mechanical Turk: Gold mine or coal mine?* Computational Linguistics (editorial), 37(2):413–420.

IDE N., BAKER C. AND PASSONNEAU R. (2008). *Masc: the manually annotated sub-corpus of american english.* In Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08), Marrakech, Morocco: European Language Resources Association (ELRA). http ://www.lrec-conf.org/proceedings/lrec2008/.

IDE N. AND MACLEOD C. (2001). *The american national corpus: A standardized resource of American English*. In Proceedings of Corpus Linguistics 2001, volume 3.

LENAT D. (1995). *CYC: A large-scale investment in knowledge infrastructure*. Communications of the ACM, 38(11):33–38, 1995.

LAFOURCADE M., LE BRUN N. AND A. JOUBERT A. (2015). *Games with a Purpose (GWAPS)* ISBN: 978-1-84821-803-1 July 2015, Wiley-ISTE, 158 p.

LAFOURCADE M., AND JOUBERT A. (2010). *Computing trees of named word usages from a crowdsourced lexical network.* Investigationes Linguisticae, vol. XXI, pp. 39-56.

LAFOURCADE M., (2007). *Making people play for Lexical Acquisition*. In Proc. SNLP 2007, 7th Symposium on Natural Language Processing. Pattaya, Thailande, 13-15 December 2007, 8 p.

LAFOURCADE, JOUBERT A., SCHWAB D., AND ZOCK M. (2011). *Évaluation et consolidation d'un réseau lexical grâce à un assistant ludique pour le " mot sur le bout de la langue "* In proc of TALN'11, Montpellier, France, 27 juin-1er juillet 2011, pp. 295-306.

LIEBERMAN H., SMITH D., AND TEETERS A. (2007). *Common Consensus: A Web-based Game for Collecting Commonsense Goals Workshop on Common Sense for Intelligent Interfaces*. ACM International Conference on Intelligent User Interfaces (IUI-07), Honolulu, January 2007.

MORO A. AND NAVIGLI R. (2015). *Semeval-2015 task 13: Multilingual all-words sense disambiguation and*

entity linking. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), p. 288–297, Denver, Colorado : Association for Computational Linguistics.

NAVILI R. & PONZETTO S. P. (2010). *Babelnet: Building a very large multilingual semantic network.* In Proceedings of the 48th annual meeting of the association for computational linguistics, p. 216–225 : Association for Computational Linguistics (ACL).

RAGANATO A., CAMACHO-COLLADOS J., AND NAVIGLI R. (2017). *Word sense disambiguation : A unified evaluation framework and empirical comparison.* In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics : Volume 1, Long Papers, p. 99–110, Valencia, Spain : Association for Computational Linguistics.

SAGOT B., FORT K., ADDA G., MARIANI J., AND LANG B. (2011). *Un turc mécanique pour les ressources linguistiques : critique de la myriadisation du travail parcellisé.*, Traitement Automatique des Langues Naturelles (TALN), Montpellier, France, 2011.

VIAL L., LECOUTEUX B. ET SCHWAB S. (2017). *Uniformisation de corpus anglais annotés en sens.* In proc of TALN 2017, 26 – 30 juin 2017, Orléans, 6 p.

VENHUIZEN N., EVANG K., BASILE V., AND BOS J. (2013). *Gamification for Word Sense Labeling.* Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013), Mar 2013, Potsdam, Germany.