# Playing Alias - efficiency for *wordnet(s)*

**Sven Aller[1], Heili Orav[1], Kadri Vare[1], Sirli Zupping[1]**

[1]University of Tartu
Liivi 2-308, Tartu, 50409, Estonia

{sven.aller, heili.orav, kadri.vare, sirli.zupping}@ut.ee

## Abstract

This paper describes an electronic variant of popular word game Alias where people have to guess words according to their associations via synonyms, opposites, hyperonyms etc. Lexical data comes from the Estonian Wordnet. The computer game Alias which draws information from Estonian Wordnet is useful at least for two reasons: it creates an opportunity to learn language through play, and it helps to evaluate and improve the quality of Estonian Wordnet.

## 1 Introduction

WordNet[1] is one of the most well-known lexico-semantic resources which is not used simply as a thesaurus for linguistic knowledge but also for language technology applications of language technology. Tony Veale has said that "WordNet … has found myriad applications in the field of natural language processing[2]" (i.e word sense disambiguation, ontologies, wordnets for opinion mining or sentiment analysis etc).

Estonian Wordnet (EstWN)[3] has grown quite large in size and our team is consistently working on the wordnet quality improvement. Since it is fairly complicated to revise concepts and their semantic relations manually (even one-by-one), automatic or semi-automatic ways for checking and discovering errors are preferred. For checking the consistency of EstWN different test patterns (Lohk 2015), also word frequency lists and corpora were used. One of the possibilities is to use gamification in language learning, namely a word explanation game called Alias. The Estoni-an computer game Alias[4] uses nouns, verbs, adjectives and adverbs present in EstWN[5]. In this paper we describe firstly how Alias is compiled and secondly, how it helps to improve the quality of EstWN. Although the data for learning language is quite useful and interesting, it is not the primary focus of this paper.

## 2 Estonian Wordnet

When setting up the Estonian WordNet we followed the principles of Princeton WordNet and EuroWordnet[6]. EstWN was built as a part of the EWN project (EuroWordNet-2 from the beginning of January 1998) and thus used the extension method as a starting point. It means that Base Concepts from English were translated into Estonian as a first basis for a monolingual extension. The extensions have been compiled manually from Estonian monolingual dictionaries and other monolingual resources (like frequency lists from Corpora of Written Estonian[7]).

EstWN includes nouns, verbs, adjectives and adverbs; as well as a set of multiword units. The database currently (September 2015; version 72) contains approximately 75 000 concepts (within more than 95 000 words) which are connected with approx 210 000 semantic relations and work is still in progress.

## 3 Design of the computer game Alias

Based on Princeton WordNet a game for word sense labeling has been created (Venhuizen et al 2013)[8]. Since obtaining gold standard data for word sense disambiguation is costly, they are using gamification for collecting semantically annotated data. Another game that uses Princeton WordNet is an on-line questions game Piclick[9].

---

[1] http://wordnet.princeton.edu
[2] http://www.odcsss.ie/node/39
[3] http://www.cl.ut.ee/ressursid/teksaurus/

[4] http://keeleressursid.ee/alias/
[5] http://www.cl.ut.ee/ressursid/teksaurus/
[6] http://www.illc.uva.nl/EuroWordNet/
[7] http://www.cl.ut.ee/korpused/
[8] http://wordrobe.housing.rug.nl/Wordrobe
[9] https://kask.eti.pg.gda.pl/pinqee/game

This is an implementation of twenty questions game, where one person thinks of a concept while the other asks him a series of yes/no questions and attempts to guess what his partner thinks of (Rzeniewicz and Szymanski, 2013).

One of the computer games which uses concepts and relations between these concepts is called word explanation game Alias, where the goal is to explain words to one's partner using different hints. These hints are typically definitions, synonyms, antonyms, hyperonyms and hyponyms etc, which are mostly present in wordnet making it suitable knowledge base for Alias' game engine.

Alias as a computer game is designed to be used by non-experts, non-linguists, and for players to play for fun. One of the main crowdsourcing platform is Amazon's Mechanical Turk, where workers get paid. In Alias game it assumed that contributors are awarded with entertainment and players are challenged to win more points than the computer.

The computer chooses a random word and shows different hints which are supposed to help a player guess the right words. For each word up to 12 randomly chosen hints are given. Hints are given to a player in sequence. If the player does not guess the word by the last hint, the point will be given to the computer.

Alias is written in PHP and it is web-based. Considering the game's architecture the EstWN database is somewhat modified – Alias uses only these synsets which have at least three hints to show (synonyms or other semantic relations), which in turn means, that at least three hints for a player are assured.

### 3.1 Different levels of Alias

The EstWN contains of words, which have very different usage frequencies and it can be quite complicated to guess the words, which are rarely used (mostly adverbs, i.e *criss-cross*) or domain-specific (i.e grammatical categories in linguistics, *ablative case*) for example. For this reason words for Alias game are selected in comparison of the word frequency lists from the Corpus of Written Estonian[10] and only these words from the synsets that belong to the frequency list are selected for playing. Following Table 1 shows the numbers of words per word classes of different levels in Alias game. Words are selected as follows: words from EstWN which are also in the

list of most frequent words, this means that conjunctives and pronouns are left out from the frequent words, since they do not exist in EstWN. Also, only one member of the synset is taken from the frequent words list, for example if both synset members are in the frequency list ('kid' and 'child') then only the first is chosen.

**Table 1.** Numbers of words of different levels in Alias game

|  | Beginner (selected from 1000 frequent words) | Intermediate (selected from 5000 frequent words) | Expert (selected from 10000 frequent words) |
|---|---|---|---|
| **Nouns** | 333 | 1654 | 2863 |
| **Verbs** | 161 | 583 | 883 |
| **Adjectives** | 56 | 315 | 528 |
| **Adverbs** | 99 | 251 | 384 |
| **All** | **649** | **2803** | **4658** |

Based on that information there are three different levels: beginner level contains of 649 words (selected from 1000 frequent), intermediate level contains of 2803 words (selected from 5000 frequent) and expert level of 4658 words (selected from 10 000 frequent). Homonyms are connected, the word *bank*, for example, displays hints from the meanings of both institution and natural object.

### 3.2 Questions for Alias

There are 55 different types of semantic relations present on Alias game (as it is in EstWN). In addition also definitions and example-sentences are used. Every type of semantic relation is related to a certain sentence template, which is presented to a player. The sentences should be simple in the sense that an average user is supposed to understand the questions that present different semantic relations.

Here are presented some of the sentence templates which Alias uses for questions:
- antonym – It's opposite for ___ (for example "*It's opposite for a man*")

- fuzzynym – It's somehow related to _____ (for example "*It's somehow related to the word elegance*")

Similarly to original board game Alias the computer game also asks words in dictionary form – nouns in nominative and verbs in infinitive form.

Estonian language is rich in compound words and in EstWN many hyponyms contain of their hyperonym as the second part of the compound word.

1. For example: one type of *kaabu* 'hat' is *vilt+kaabu* 'trilby hat'

If the compound word consists of the word that is currently guessed, the similar stems of the words are removed (see example 2). The same rule applies also in the original board game. Since Estonian is rich in cases, persons and in inflectional system, then it is quite complicated to find the word with the similar stem. The morphological analyzer [11] is used to compare the lemmas in hint to the lemma of the asked word. If they match, then the similar stem is replaced with a gap.

2. For example:
Question:
See on teatud liiki õunapuu.
*This has a type of appletree.*
is replaced
See on teatud tüüpi õuna_____
*This has a type of apple_____*
Answer: Puu (*Tree*)

Question:
You can use this word like that:
*Bring back my pony to me*
is replaced with
*Bring _____ my pony to me*
Answer: Back

## 4   Some statistics from play log

Since the December 2014 Alias is played 664 times. During these games, 2571 words have been asked, it means that average 3,87 words per game are guessed. As the Table 2 shows, the correctly guessed words percentage differed largely across different semantic relations and definitions or examples used.

All the semantic relations present in EstWN are also used in Alias. Of course there are some

relations in EstWN, which are not so frequent – role_instrument or has_mero_member for example, which means that they are also asked less frequently during the game. Table 2 states that the top-guessed relation is role_instrument even though it occurred only 5 times, so we can say that it is not statistically so important as definitions and antonym relation for example.

Groups (as group_role, group_xpos, group_holo, group_involved, group_derive) are connected in table because they share the same sentence template for hints. These sentence templates will be changed in the next version of the game.

## 5   Discussion

George Miller, as a psycholinguist was interested in how the human semantic memory is organized (Miller 1998), which type of relations are most typical between words and concepts.

In addition to (psycho)linguistic tests, some conclusions/inferences can be drawn using log files of game Alias as well. Results give us feedback which relations are clear, which are too fuzzy or too general or just too strange. For example: *migration* involved_location *residence, abode*. Piek Vossen's (2002) test for location_involved relation is:
*(A/an) X is the place where the Y happens.*
So, it is obvious that relation between *migration* and *residence* needs to be corrected in EstWN.

As you can see from the Table 2, there is a slight difference between guessing hints containing of hyperonyms (7.2%) and hyponyms (9.1%), the latter shows slightly better results. Hyperonyms might be too general, they might have multiple hyponyms, for example 'to run – to move'. While giving a hyponym as hint, for example 'to run – to sprint', opens the meaning of the word more precisely.

Since fuzzynym-hints do not appear to be very useful for players (only 7.1%), we can assume, that the connections and associations presented by fuzzynyms are too vague. Some of the fuzzynynms can be assigned to a more specific semantic relation, for example 'doctor' and 'stetoscope' or 'postman' and 'postbag' which denote something that belongs to some certain profession. But, as we could see from the play logs, there are many fuzzynyms completely distant, for example 'presentation' and 'evolution', 'painting' and 'education' etc.

From the player's perspective the definitions (21.3%) and examples (18.2%) are one of the

---

[11] http://www.filosoft.ee/html_morf_et/

most successful hint for guessing the right word. In many cases we can see from logs that various hints with semantic relations do not help the player, but definition and explanation – also even if they are the first hints – are very informative. This means that as a concept based database EstWN needs to have clear definitions and good examples to open the meanings of concepts.

The meaning of the word is quite well guessed while hints present synonyms (here Variants, 14.5% right answers) or antonyms (33.7%) and near antonyms (9.0%) or near synonyms (9.4%). It is intuitively simpler to guess for example the word 'kiss' by its synonym 'buss' than its hyperonym 'touch' or verb 'to buy' by its antonym 'to sell' than its hyperonym 'to acquire'.

Hints that contain of functional relations (i.e role, meronymy) are usually very clear to a player, of course these indicate to concrete objects. The role-relation can connect both nouns to nouns and nouns to verbs. For example the verb 'to run' has been guessed by its role_agent 'runner' but not by its hyperonym 'to move'.

The logs from beginner and even intermediate level can indicate to problems of the main vocabulary, for example for a question: this is near synonym for the word 'swamp bridge' the correct answer should be 'road'. Of course this near synonym link is not correct and should be revised also in EstWN.

In many aspects this game reflects that the associations of words/concepts are free and arbitrary in human minds. For example, illegible (sloppy, quickly written) handwriting can remind us the doctors' style of handwriting. But still it is possible – if considered carefully and thoroughly – find a certain system, which is similar to the one Georg Miller started to create a model of the human mental lexicon. In „On wordnets and relations" (Piasecki et al 2013) is mentioned that forming a synset (in the sense of wordnet) is a quite difficult task and has been largely left to the intuition of people who build wordnets. Game gives us a chance to check how similar the compilers intuition is to a player's intuition.

## 6   Conclusion

The play logs contain of valuable information for a lexicographer and using this for improvement of EstWN is quite a new approach. The EstWN has benefited from the Alias game in many ways. Firstly it was possible to determine completely false synsets and/or the non-suitable semantic relations. Secondly it was possible to correct some of the semantic relations. Thirdly some of the definitions were improved and made more precise. The correction work has grown more systematic, since more log files have become available. As an addition to revising and correcting synsets and their relations it was interesting to observe which hints were more informative to players than the others. It gives us good feedback if there is any semantic relation too general, too narrow or just too vague.

Not less important is the value to Alias game and it working principles. If studying the logs more thoroughly it is possible to improve the quality of Alias, for example how to choose concepts, how to sort, choose, form and present hints etc. This game is adjustable for every language which has their own wordnet.

Researchers of Polish Wordnet (Maziarz et al 2013) have said that "Synonymy is intended as the cornerstone of a wordnet, hypernymy – its backbone, meronymy – its essential glue". After analyzed the log files of Alias-game we can say that traditional definitions and antonyms are clearer to a player with no linguistic background.

## References

Fellbaum, Ch., 2010. WordNet, in: Poli, R., Healy, M., Kameas, A. (Eds.), Theory and Applications of Ontology: Computer Applications. Springer Netherlands, pp. 231–243.

Lohk, A. 2015. A System of Test Patterns to Check and Validate the Semantic Hierarchies of Wordnet-type Dictionaries. Thesis on Informatics and System Engineering C105. Press of Tallinn University of Technology.

Maziarz, M., Piasecki, M.; Szpakowicz, S. 2013. The chicken-and-egg problem in wordnet design: synonymy, synsets and constitutive relations. – Language Resources and Evaluation 47 (3), 769–796.

Miller, G. A. 1998. Nouns in WordNet. – WordNet. An Electronic Lexical Database. Ed. Christiane Fellbaum. Cambridge, MA: The MIT Press, 23–46.

Piasecki, M.; Szpakowicz, S.; Fellbaum, Ch.; Pedersen, B.S. 2013. On wordnets and relations. – Language Resources and Evaluation 47 (3), 757–767.

Rzeniewicz, J.; Szymański, J. 2013. Bringing Common Sense To Wordnet With A Word Game. *Computational Collective Intelligence. Technologies and Applications* (2013): 296-305. Web. 14 Sept. 2015.

Venhuizen, N.J., Basile, V., Evang, K. and Bos, J. 2013. Gamification for Word Sense Labeling. In Katrin Erk and Alexander Koller (eds.), *Proceed-*

*ings of the 10th International Conference on Computational Semantics (IWCS'13) -- Short Papers*, 397-403, Potsdam, Germany.

Vossen, P. 2002. EuroWordNet General Document. Version 3. Final. July 1, 2002. http://www.vossen.info/docs/2002/EWNGeneral.pdf (27.10.2015).

**Table 2.** Results of playing by different relations

| Relation | Occurence | Right cases | Right cases (%) | Wrong cases |
|---|---|---|---|---|
| role_instrument | 5 | 3 | 60.0% | 2 |
| role_agent | 17 | 7 | 41.2% | 10 |
| antonym | 86 | 29 | 33.7% | 57 |
| causes | 18 | 6 | 33.3% | 12 |
| has_holo_madeof | 23 | 6 | 26.1% | 17 |
| DEFINITION | 1390 | 296 | 21.3% | 1094 |
| is_caused_by | 31 | 6 | 19.4% | 25 |
| EXAMPLE | 1136 | 207 | 18.2% | 929 |
| **group_role** | 41 | 6 | 14.6% | 35 |
| VARIANTS | 1597 | 232 | 14.5% | 1365 |
| has_mero_member | 7 | 1 | 14.3% | 6 |
| has_mero_madeof | 7 | 1 | 14.3% | 6 |
| has_meronym | 26 | 3 | 11.5% | 23 |
| has_mero_part | 36 | 4 | 11.1% | 32 |
| has_holo_member | 18 | 2 | 11.1% | 16 |
| **group_involved** | 42 | 4 | 9.5% | 38 |
| near_synonym | 577 | 54 | 9.4% | 523 |
| has_hyponym | 2123 | 194 | 9.1% | 1929 |
| near_antonym | 200 | 18 | 9.0% | 182 |
| **group_holo** | 60 | 5 | 8.3% | 55 |
| has_mero_location | 12 | 1 | 8.3% | 11 |
| role_location | 13 | 1 | 7.7% | 12 |
| has_hyperonym | 994 | 72 | 7.2% | 922 |
| has_xpos_hyponym | 152 | 11 | 7.2% | 141 |
| fuzzynym | 622 | 44 | 7.1% | 578 |
| **group_xpos** | 313 | 19 | 6.1% | 294 |
| state_of | 84 | 4 | 4.8% | 80 |

| | | | | |
|---|---|---|---|---|
| be_in_state | 45 | 1 | 2.2% | 44 |
| is_subevent_of | 4 | 0 | 0.0% | 4 |
| has_mero_portion | 2 | 0 | 0.0% | 2 |
| has_holo_portion | 2 | 0 | 0.0% | 2 |
| role_target_direction | 1 | 0 | 0.0% | 1 |
| has_subevent | 1 | 0 | 0.0% | 1 |
| role_manner | 1 | 0 | 0.0% | 1 |
| has_holo_location | 0 | 0 | 0.0% | 0 |
| belongs_to_class | 0 | 0 | 0.0% | 0 |
| **group_derive** | 0 | 0 | 0.0% | 0 |
| role_source_direction | 0 | 0 | 0.0% | 0 |
| has_instance | 0 | 0 | 0.0% | 0 |
| role_direction | 0 | 0 | 0.0% | 0 |