IEEE Second International Conference on Data Stream Mining & Processing
August 21-25, 2018, Lviv, Ukraine

# Fuzzy Reconstructions in Linguistics

Olena Basalkevych
*Medical Informatics Department*
*Danylo Halytsky Lviv National Medical University*
Lviv, Ukraine
olena.basalkevych@gmail.com

Olexandr Basalkevych
*Product Engineering and Development*
*Globallogic Inc.*
Lviv, Ukraine
basalkevych.alex@gmail.com

***Abstract* — Application of fuzzy modeling in historical linguistics constitutes the contribution of the paper. Synonymic sets and word semantic structures of an Older Scots adjective are interpreted as associative fields, and on the other hand, as fuzzy sets. The resulting fuzzy associative linguistic diachronic reconstructions, including fuzzy associative fields (FAF) and fuzzy associative word structures (FAWS), along with their combination into a fuzzy associative adjectival network (FAAN) with further fuzzy semantic differentiation and lexis stratification, represent a fragment of a fuzzy adjectival associative thesaurus of Older Scots. Generally saying, the work demonstrates the use of artificial intelligence tools in diachronic linguistic, which is rather unexperienced practice.**

***Keywords* — fuzzy associative field, fuzzy associative word structure; fuzzy associative adjectival network, fuzzy diachronic semantic differential, fuzzy lexis stratification, fuzzy adjectival associative thesaurus of Older Scots.***

## I. INTRODUCTION

Vague limits of linguistic phenomena promise a good success to the adaptation of fuzzy logic to the area. Lotfi Zadeh, the honoured father of fuzzy logic, argued that probability lacks sufficient “expressiveness” to deal with uncertainty in the natural language [1]. The specially prolific ground for fuzzy upgrading seems to be a linguistic reconstruction, the main research tool of the historical linguistics. In the concrete case, Older Scots lexicography makes the language canvas for the modeling. Synonymic sets and word semantic structures of an Older Scots adjective find their associative interpretation in neurolinguistics, and then, fuzzy evaluation in fuzzy logic. The consequent fuzzy associative linguistic diachronic reconstructions do not have any precedents, their synchronic prototypes are G.Kiss’ probabilistic associative thesaurus [2], Osgood's semantic differential and Zadeh’s fuzzy stratification [14]. Supported by the Medical Research Council of Great Britain, the associative experiment issued into a weighed graph of G.Kiss, recreating the associative habits of the speakers of the English language, their cognitive and mental dispositions carved in the language is a pharos of the research strategy. Osgood's Semantic Differential represents a rating scale designed by Osgood to measure the semantics or meaning of words, particularly adjectives, and their referent concepts. The last lecture of Lotfi Zadeh represents a fuzzy categorization instrument – stratification of reference information [14]. All these works contributed greatly to the evolved mathematical model. The reconstructions of the sort in diachronic linguistics are not revealed, although the great computational steps were made since the early seventies by Eastlack, Burton-Hunter, Remmel, Hewson, Kondrak and others [3] in identifying cognates, reconstructing proto-forms, deriving reflexes, generally saying, in equipping the theory of comparative method with robust reconstructive phonology instruments all based on the manipulation with recurrent sound correspondences of phonemes in cognates.

## II. PRE-RESEARCH

### A. Diachronic synonymic sets of Older Scots

The initial point of the study was the construction of Older Scots adjectival synonymic strings on the basis of the Dictionary of the Older Scottish Tongue (DOST), a part of the Dictionary of the Scots Language, available online [13]. The method involved the words from the English synonymic sets as subjects of the Dictionary surfing. The DOST answers formed unexpectedly large strings of semantically close words of very different connotations. Any word of the string was weighted following the formula:

$$W_i^{(y)} = \frac{y_i - y_{\min}}{y_{\max} - y_{\min}}\left(\frac{1}{n+1} - 1\right) + 1 \qquad (1)$$

where ( $y_{\max} - y_{\min}$ ) is a diachronic range of a set, ( $y_i - y_{\min}$ ) is a distance to the appearance of the concrete word in a set, and n – the number of words in a set [4]. The words were ordered in the way of ascending their advent dates so that the oldest ones were the first in the string. Any word was equipped with advent year, weight coefficient, orthography forms, etymology, meaning, diachronic text prototype, its literature source and author. The Access database IsetIntro was configured for the purpose [5].

### B. Diachronic semantic word structure

On the basis of the mentioned adjectival sets, the database query for registering word entries into different sets was organized in order to detect all available meanings of a word, together with a text prototype, its author and date that is a criterion for the structure expanding. The entities were called diachronic semantic word structures [5].

## III. ASSOCIATIVE INTERPRETATION

The multi-coloration of connotative values of set words and the presence of precedent texts makes it possible to reflect the study into associative linguistics’ plane. The mentioned diachronic synonymic sets of Older Scots could be interpreted as associative fields whilst DOST would be treated as a collective Older Scots brain representing its answers to the stimuli within the framework of an associative experiment (AE) with “historical reaction”. The principle of AE says that the first replicas are the strongest associations based on the most frequent usage.[6] Meanwhile the Zipf’s law states that the earliest words are the most frequently used ones [7]. Since the first positions in the gained synonymic sets belong to the oldest words with the biggest weight coefficients, according to the Zipf's law we have kept the norm of AE and by this received the measure of associativity – the weight coefficient $W_i$. Consequently, the diachronic semantic word structure could also gain its associative interpretation. Having turned the diachronic component $W_i$ into an associativity one, the main quantitative formant of a semantic structure becomes the measure of similarity or associativity of a certain word with different semantic features. Therefore it could be called an associative word structure.

## IV. FUZZIFICATION

Associativity and vagueness seem to have a feature in common. One thing could be associated with others due to different factors (similarity, relevance or opposition) and just to some certain extend. In our case, the words of an associative field are associated with a forming semantic feature due to their synonymic nature or similarity that is rarely going to be absolute. The level of appropriateness or significance of set's components and the degree of belief or confidence (trueness) level of the made statements in fuzzy logic is introduced by a membership function (MF) [8]. The coefficient $W_i$ favorably fits the nature of MF, introducing the level of word associativity with a semantic feature within the range [0;1]. So the words with higher MFs are treated to be more characteristic in the set and better associated with the semantic feature regarded. Therefore the fuzzy version of a mathematical model whose role is to regulate and prognosticate the studied process is expected to be the most profitable one.

### A. *Fuzzy associative fields*

The mathematical model of research considers an associative field as a fuzzy subset of the set of all language adjectives. The subset is named by a semantic feature (a dominant) grouped under with elements allotted by MFs calculated according to formula 1. The fuzzy associative field (FAF) could be illustrated by e.g. Lucky={happy/1, hapin/0.99, sely/0.93, wel/0.9, wele/0.9, fortunit/0.76, fortunate/0.63, mervailous/0.63, fortunable/0.6, chancy/0.59, ewrous/0.53, lukkie/0.51, sonsy/0.4, canny/0.07}

A semantic feature or a dominant plays here the role of a linguistic variable, and all indicated subset names constitute its term-set: Dominant={Flexible, Inflexible, Brave, Cowardly, Intelligent, Stupid, Large, Small, Lucky, Unlucky,… }[9].

### B. *Fuzzy associative word structure.*

Taking through the same fuzzification of the coefficient $W_i$, the associative word structure becomes a fuzzy one. It is clear that a word could be associated with this or that meaning to different extend, being closer to its direct value, and farther from a lateral one. The example for the word *Happy* is derived from the represented above FAF *Lucky*, together with these of *Successful*, and *Competent*:

Happy= {lucky/1, successful/0.57, competent/0.38}

First of all the structure is extended for the element *competent* as compared to [8], since the database has been enriched with more entries so far. Analyzing the MFs, we should state that *Happy* is the most associated with the semantic feature *lucky* (or belongs mainly to the fuzzy subset *lucky*, in terms of fuzzy logic), less associated with *successful* and the least associated with *competent*.

Hence, we will call the structure of this sort a fuzzy associative word structure (FAWS) [8].

### C. *Fuzzy associative adjectival network. N-association trees. Proper application software.*

The articulated two categories of words: meanings associated with a word, and words rendering the same value are the main relations in G.Kiss' Associative Thesaurus [3]. So, combining FAFs through their elements' FAWS results into a fuzzy associative adjectival network (FAAN). The example of it, connecting fields *Good, Strong, Inflexible, Lucky* and *Successful* through numerous FAWS including *Happy* is represented in fig. 1 [9].

The mentioned lexicon can be easily animated by the brain function of the speaker. To simulate the process, we are activating the created fuzzy network, where formally we can discern the set of places P, the set of transitions T, the input function I and the output function O. The input function I reflects the transition $t_j$ into the set of positions I($t_j$) called the initial positions of a transition release. The output function

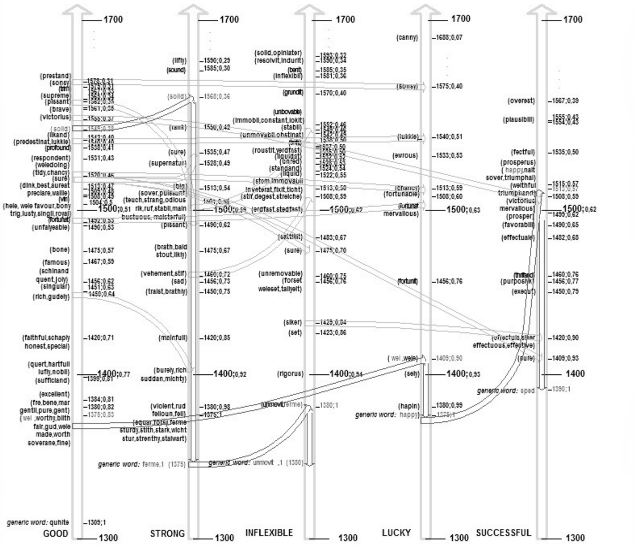
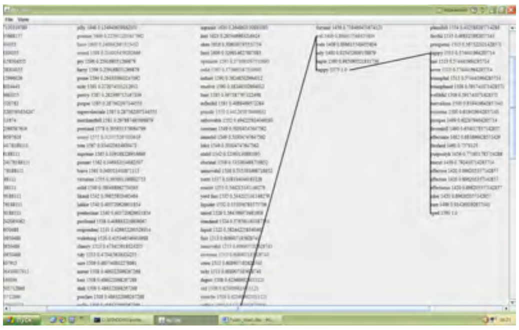



Fig. 1. Fragment of FAAN.

O reflects $t_j$ into the set of output transition positions O($t_j$). Fuzziness is introduced into the set of places P, taken by words with appropriate MFs. The release of a transition rearranges the fuzziness distribution of the route, and the output place-word according to the rule $\mu(x_1 \cap x_2) = \min[\mu(x_1), \mu(x_2)]$ gains the minimum value of MFs of the input positions [9]. The fruit of this animation is in the probing of the n - step environment of any word within the net, the technique introduced by Kiss in his weighed graph. According to the FAWS of the regarded word, the transition to the most significant (with a maximum MF) output position which belongs to another associative field will be released. The following transition will be made to a dominant of the reached field. For the dominant we repeat the same algorithm. At the 2n-th step we will reach n-association to the studied stimulus since as was already mentioned two transitions normally lead to any new association gain: between-field one and inter-field one. The performance of the constructed fuzzy associative adjectival network is oriented to the maximum belief (maximum MF) but is still sensitive to the weakest link of the route [9].

The mentioned algorithm of association-chain building is implemented with Java. The library Swing is involved for data visualization. The architecture is based on MVP design template (*Model-View-Presenter*) that delivers the visual reflection and event-procession behaviour into different classes, namely *View* and *Presenter*. The input data are given into format XML. The model DOM (*Document Object Model*) is employed for the work with XML-format. First of all the file of the model should be chosen and downloaded into the program. Then a user is pressing any word with the help of a mouse. The program visualizes the connections among words and calculates the general chain fuzziness. For example, association chain for *Victorious* is realized in Fig.2a. It is *Victorious–Sped* with MF or a belief degree 0.37. Associative reaction to *Wel* studied manually in [9] is *Happy–Sped* with belief 0.57 (Fig.2b). The association chain *Solid-Firm* is reached with the belief degree 0.36 (Fig.2c).


a


b

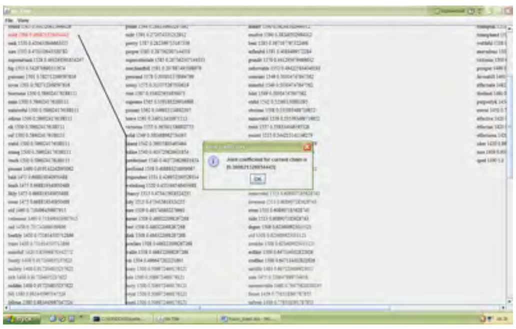

c

Fig. 2. Program realisation of association chains: a) Victorious–Sped; b) Happy–Sped; c) Solid-Firm.

*D. Two way FAAN*

The following research step is an addition of an antonymous flank to the received structure. It is known from AE that the largest associative power belongs to the antonym of the stimulus. So we connect the antonymic fuzzy associative field to the original one, creating the base of a triangular plane. The basic opposition is the pair of a dominant and its antonym: Successful-Unsuccessful, Lucky-Unlucky, Flexible-Inflexible, Strong-Weak, Good-Bad with two way connection (Fig.3). In the case, an association chain is going to contain two branches: positive reactions from Successful-Lucky-Inflexible-Strong-Good and negative from Unsuccessful-Unlucky-Flexible-Weak-Bad, thus creating an association tree. For instance, we can choose for the stimulus the regarded in *IVc* word Solid. Evolving the algorithm of network simulation, the other 2n steps are applied to the time-antonym of the studied stimulus with a little bit shuffling order: the transition to a dominant is realized immediately since the gravity to the main bearer of the meaning triggers first [8].

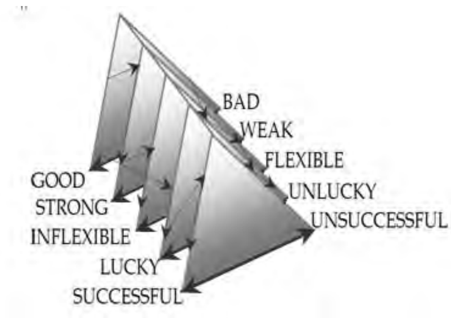



Fig. 3. The model of two-way FAAN.

Then we get the enriched chain reaction Solid=(+)Ferme/0.36And(–)Lidder-Wikit-Perilous/0.38. The general association will get 0.36(=min (0.36;0.38)) belief degree, which is not too strong sureness that the reaction of a Scottish man of 1300-1700 to the word *Solid* would be like this.

*E. Fuzzy diachronic semantic differential*

The special interest is attracted to the triangular cut of the structure. If to unbend the wings of the figure, the chrono-model of Osgood's semantic differential could be clearly observed. Following the same Zipf's law background along with fuzzy interpretation of the weight coefficient $W_i$, we proceed to fuzzy diachronic semantic differential (FDSD). Its normalized variant is shown in Fig.4.

All the meaning subtleties for concepts' dichotomy Strong-Weak are introduced along the triangle sides. The chaotic layout of the meanings is normalized by the introduction of time intervals [11]. Now with the help of this historic evaluative device we can estimate the main characteristics of certain words throughout the history of Older Scots. Following the represented in DOST diachronic text prototypes of the reporting words of our differential, we can state that:

**Companye** could be **buyrly** with the degree of belief **0.917.**

**Women** could be **odious** with the degree of belief **0.586**

**Watteris, fluidis** could be **proud** with the degree of belief **0.255**

**Leggs** could be **wankle** with the degree of belief **0.019**

**Me** could be **bauche** with the degree of belief **0.290**

**Complexioun** could be **selie** with the degree of belief **0.606**

**Woman** could be **brukil** with the degree of belief **0.984**

The most prominent representatives of syntagmatic associations are taken from more numerous selection of [12]. The short research issue reveals that according to Older Scots pattern of cross-cultural universal Strong-Weak the negative personality descriptor for **woman** is more often **brukil** than **odious**: Woman={brukil/0.984, odious/0.586}.

By this study we are extending our reconstruction with syntagmatic associations, the very important component of Kiss's Associative Thesaurus and AE in general.

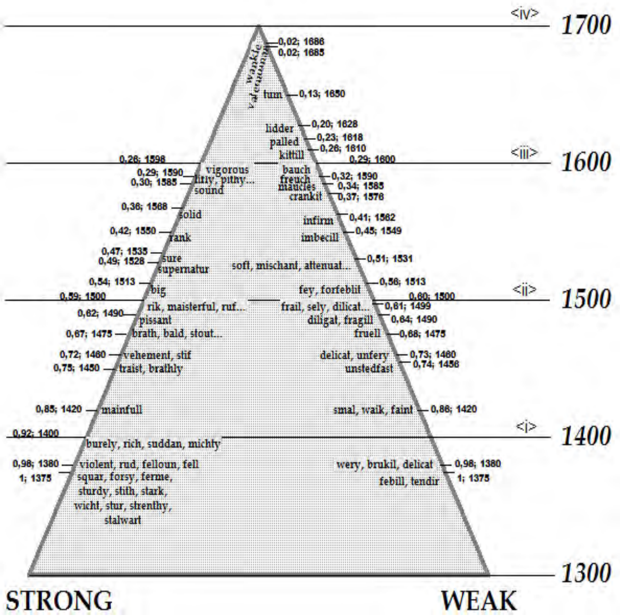


Fig. 4. Model of the normalized FDSD.

The complete picture claims to be the fragmental fuzzy adjectival associative thesaurus of Older Scots, and the monads of words collocated with a DOST reaction-words should enrich the previous research practices. The propagation of fuzziness through the fuzzy network is playing here its activating role.

The other known function of a semantic differential is to evaluate the psycho-linguistic portrait of a person. In a concrete case we could speculate about characteristics of a collective Older Scottish speaker represented by the myriad of Scottish penmen. For instance, the mentioned fuzzy subset Woman reveals the expressions of a negative attitude of an average Older Scot towards females. The aforesaid makes the sketch of prospect semantic research following all the conventional cross-cultural universals (evaluation, potency and activity): Good-Bad, Strong-Weak, Active-Passive.

### *F. Fuzzy lexis stratification*

The scheme in Fig.2 implies another research-strategy – the stratification of lexis. The time intervals, common for both flanks, comprise the time lexis oppositions, combining them into four strata. The idea of fuzzy stratification belongs to late Lotfi Zadeh, and is widely involved into processing encyclopedias and notebooks. In the concrete case, the fuzzy analysis could open possibilities of ascertaining the lexis stratum gravity poles, time antonymy and significance status of the strata, all enabled by the manipulation with border MFs. The fuzzy variable Gravity = (strong/ $\mu_1(x)$; weak/ $\mu_2(x)$) is introduced in any stratum, where $\mu_1(x)$ and $\mu_2(x)$ are the MFs of the sectional antonymic pair. It makes possible to evaluate the preferring concept for the time-stratum, in other words, the intensity of positive and negative concepts development during some period. Time antonymy could be analyzed not only for sectional pairs but at any moment within any stratum by straight bridging two flanks and their immediate representatives. The notion of a core and a periphery of a fuzzy set will help us to analyze the significance of a concrete stratum establishing the specific fuzziness that divides the core strata from peripheral ones, by this restricting the research concentration [11].

In our case the contrasted pairs are

1400, (Burely/0.917) - 1380, (Wery/0.984)

1500, (Rik/0.586) – 1500, (Frail/0.606)

1600, (Proud/0.255) – 1600, (Bauch/0.290)

For the taken error of 50 years, in 1700 the time opposition is absent (Fig.3).

Now adapting the fuzzy variable Gravity = (strong/ $\mu_1(x)$; weak/ $\mu_2(x)$) for any of strata, again considering the sectional MFs, we could characterize them:

1300-1400 stratum: Gravity = (strong/0.917; weak/0.984);

1400-1500 stratum: Gravity = (strong/0.586; weak/0.606);

1500-1600 stratum: Gravity = (strong/0.255; weak/0.290).

The Gravity gives arguments in favour of the Weak concept for all strata. So negative, a more developed since more exciting and alarming concept, becomes the gravity pole for the whole stratification figure [12].

For the core strata let us adopt MF 0.25 as a fuzzy measure. Assimilating this level for both flanks we cut the last 1600-1700 stratum and analyze the preceding three only.

## V. CONCLUSIONS

New notions of fuzzy associative fields and fuzzy associative word structure benefit from the realized symbiotic approach. Being combined together, they constitute the fuzzy associative adjectival network, a singular vertical cut of which is an extended fuzzy modification of a semantic differential or a fuzzy diachronic semantic differential. Another evaluation attitude, more volumetric one, considers the gained plane as a fuzzy lexis stratification structure consisting of the strata of lexical oppositions. The described metamorphoses are realized with the help of the database in Access and implemented with Java using Eclipse IDE. The prospect of the research is the evolution of fuzzy adjectival associative thesaurus of Older Scots. The practical implementation of the proposed fuzzy verbal network opens possibility of simulating brain activity of the speakers of more or less antique languages revealing their cognitive and mental peculiarities, by this recreating and contrasting the world verbal pictures of different older languages, which is of interest for historical cognitivists.

## ACKNOWLEDGMENT

The special gratitude should be expressed to Vitaliy Arseniyovych Lishchynsky (1940-2015), the supervisor of the author's diploma project "Fuzzy network quality-research and modeling" performed in 1997 in "Lviv Polytechnic" State University at the Department of Applied Mathematics, whose advanced ideas and benevolent assistance made the premises for evolving the proposed mathematical model.

## REFERENCES

[1] M. Laviolette, J. W. Seaman, Jr., J. Douglas Barrett, and W. H. Woodall, "A Probabilistic and Statistical View of Fuzzy Methods," Technometrix, Taylor & Francis, Ltd., Vol. 37, No. 3, pp. 249-261, Aug., 1995.

[2] G. R. Kiss., C. Armstrong, R. Milroy, and J. Piper, "An associative thesaurus of English and its computer analysis," In Aitken, A.J., Bailey, R.W. and Hamilton-Smith, N. (Eds.), The Computer and Literary Studies. Edinburgh: University Press., 1973.

[3] G. Kondrak "Algorithss for Language Reconstruction," University of Toronto, pp.169, 2002.

[4] M. Bilynskyi, "Synonymous string as a diachronic reconstruction: The OED Middle English Evidence for verbs and deverbal coinages," Inozemna Philologia, Lviv, Vol.121, pp. 20-43, 2009.

[5] O. Basalkevych, "Diachronic model of adjective synonymic set evolution in Older Scots," IX International scientific practical conference: "Language-Culture-Personality", Scientific records of "Ostroh Academy" National University, Ostroh, Ukraine, vol. 54, pp. 91-97. 2015.

[6] M. Arapov, "Mathematical methods in hstorical linguistics," Moscow, "Nauka", 1974.

[7] D. Nelson, C. McEvoy, and S. Dennis, "What is and what does free association measure?" Free Association, pp.3-46, 2002.

[8] M. Siavavko, "Informational system "Fuzzy expert"," Lviv, Ivan Franko Lviv National University Press, 2007.

[9] O. Basalkevych, "Fuzzy modelling of associative maps," IV International scientiic practical conference "Linguocognitive and sociocultural aspects of communication," Scientific records of "Ostroh Academy" National University, Ostroh, Ukraine, vol. 56, pp. 33-38. 2015.

[10] V. Lishchynsky, and O. Basalkevych, "Anlysis of Petri Nets with one fuzzy element," the Proceedings of the Seventh Pan-Ukrainian Scientific Conference, Lviv, Ukraine, pp.61-62, 2000.

[11] O. Basalkevych, "Fuzzy diachronic differential," Proceedings of the IV Ukranian-German conference "Informatics. Culture. Technics.", Odesa, Ukraine, 30 June-2 July, pp.106-108 2016.

[12] O. Basalkevych, "A model of the diachronic semantic differential: stratification of lexis on fuzzy principles (the case of Older Scots adjectives)," Inozemna Philologia, LIFNU Press, vol. 130, pp. 16–27, 2017.

[13] "The Dictionary of the Scots Language," Edinburgh: Scottish Language Dictionaries, 2007. www.dsl.ac.ukH

[14] L. Zadeh, "Stratification, target set reachability and incremental enlargement principle," The Design of Robotics and Embedded systems, Analysis and Modeling Seminar, 250 Saturdja Dai Hall, UC Berkeley, 2016. www.youtube.com/watch?v=Ok0vJ0uykCA&list=PL8092C965854FA5BB&index=30