Analysis of functional similarities of Finnish verbs using the self-organizing map

Krista Lagus
Helsinki University of Technology
Neural Networks Research Centre
P.O.Box 5400 (Konemiehentie 2)
FIN-02015 HUT, Finland
Krista.Lagus@hut.fi

Anu Airola
University of Helsinki
Department of General Linguistics
P.O.Box 9 (Siltavuorenpenger 20 A)
FIN-00014 University of Helsinki, Finland
Anu.Airola@helsinki.fi

Abstract

Obtaining semantic or functional word categories from data in an unsupervised manner is a problem motivated both from the linguistic point of view and from that of construing language models for various language processing tasks. In this work, we use the Self-Organizing Map algorithm to visualize and cluster common Finnish verbs based on their immediate morphological contexts. Based on a data set of over 500,000 utterances of 25 verbs, we studied (1) the base forms and (2) the most common word forms of the same verbs (4764 forms). The results show that even the simple feature selection used in this experiment was found to be suitable for rough automatic categorization of verbs on the basis of data extracted from unrestricted texts. In particular, the automatically obtained organization resembles the semantic classification of verbs designed by a linguist.

1 Introduction

The data now available in the form of massive computer readable corpora has the potential to present a challenge to the way we think about the concept of linguistic theory. Especially, in theoretical corpus linguistics, one sense of the term 'data-driven' is to find new ways of deriving the theory from the data and thus bring new insights into linguistic theory itself (cf. (Sinclair, 1991; Stubbs, 1996)). However, one needs proper tools for identifying dependencies within the data and for representing them in an understandable form.

A much debated question is to what degree and which parts of our language ability are innate, and what and how much is learned from exposure to natural language in communicative settings. From the point of view of building systems capable of communication, a central question is: What is a minimal, sufficiently strong and covering set of learning principles and prior information that enables a system over time to learn versatile communicative skills? Obviously, this question defines a broad agenda for long-term research.

In this paper we discuss a sub-topic of *lexical acquisition*, i.e. the acquisition of information of semantic similarities of words (see e.g. (Boguraev and Pustejovsky, 1996)). The meaning-bearing elements in language expressions are constructions consisting of the following elements: the words themselves, morphological endings and markers, the order of words, and the intonation (see e.g. (Goldberg, 1995; Tomasello, 1998; Croft, 1999)). In actual expressions there is redundancy in the sense that several elements may contribute to communicating similar content. For example, in the sentence *I went to the store* both *went* and *to* appear to signal the presence of movement. Such redundancy allows one to deduce similarities between elements based on the contexts of their usage; in other words, to form 'contextual representations' of words based on hearing or seeing the words used in a variety of linguistic contexts (Miller and Charles, 1991; Leacock et al., 1996).

The crucial question then is what kind of information is relevant when a system forms these contextual representations. More specifically, the following question is addressed here: How much information regarding semantic or functional similarities of verbs can, in principle, be uncovered by assuming very little, i.e., by selecting the verbs to be studied based on frequency, and by looking only at their immediate contexts in a very simple and straightforward manner. It is also assumed that (1) the system is already capable of distinguishing verbs from other syntactic word classes, and (2) able to recognize some morphological features.[1]

1.1 Statistical language modeling

Statistical language models (SLM) are models that are estimated based on large corpora and that allow one to make predictions regarding future linguistic data (e.g., future words) or interpretations of such data (e.g., which word was just heard) given some properties of previous data (e.g., previous words). Such models are applied, for example, in speech recognition. Surprising success has been achieved with very simple models, such as N-gram models (cf. (Manning and Schütze, 1999)) that record frequencies of all encountered combinations of n words and attempt to predict the next word based on the previous $n - 1$ ones. However, these typically require vast amounts of learning data, and even then, some quite ordinary word combinations have not appeared in the data even once (Charniak, 1993). The central question then becomes one of generalizing existing information to new, similar but not identical cases. One way of doing this is by *clustering* the data, i.e., by grouping similar items together, and generalizing between words based on the clusters. This is sometimes called model smoothing.

A typical approach in statistical language modeling has been to apply a simple, equivalent strategy for all words, regardless of whether the words describe actions or objects. However, due to the apparent structure inherent in natural languages, a more fruitful direction might be to construct models with more structure, including different submodels for different kinds of words.

1.2 Related work

It has been previously shown that in languages with strict word order, such as English, the distribution of words in the immediate context of a target word contains a considerable amount of information regarding the syntactic and semantic category of the target word (see (Finch and Chater, 1992; Zavrel and Veenstra, 1995); for an overview of early work, cf. (Charniak, 1993)). Also the Self-Organizing Map (SOM) algorithm (cf. Sec. 2) has been applied for clustering English word forms based on the word forms in their immediate contexts in (Ritter and Kohonen, 1989; Honkela et al., 1995; Honkela, 1997), with promising results: both syntactic and semantic organization was observed. Furthermore, the word clusters obtained in such a manner have been used for encoding the meaning of documents e.g., in (Kaski et al., 1998). However, for Finnish the same approach does not seem feasible for two reasons. First, the kind of functional information that in English is coded by independent function words such as “to” and “at” is in Finnish coded by inflectional morphology (see Sec. 3 for details). Secondly, the function of word order in Finnish is to organize the flow of information at sentence and textual levels.

In models that are based on calculating combinations of word forms, the doubling of the number of target words simultaneously doubles the number of features. Moreover, with Finnish the effect is much stronger than a doubling—a single base form of a verb or a noun may have thousands of inflected forms (Karlsson, 1983). This causes immediate problems for reliable estimation of such models.

From the point of view of SLM the present article is a pilot study on how one might overcome some of the aforementioned problems. Our aim was to obtain a semantic or functional clustering for Finnish verbs. In particular, we apply the SOM algorithm for simultaneous clustering and visualization of the data. The visualization provides a means to understand and to qualitatively evaluate the obtained clustering and the feature

[1] Morphological features are often regarded as “surface features” since they are to some degree apparent in the actual words.

selection.

1.3 Structure of the article.

In Sec. 2 we briefly describe the SOM algorithm. The selected verbs and features are introduced in Sec. 3. The experiments and their results are described in Sec. 4, with conclusions and discussion in Sec. 5.

2 Self-organizing map algorithm

The Self-Organizing Map (SOM) (Kohonen, 1982; Kohonen, 1995; Kohonen et al., 1996) is an unsupervised neural network (Haykin, 1999) algorithm that is able to arrange complex and high-dimensional data so that similar inputs are, in general, found near each other on the map. On a visual display of the organized map properties of the data set can be illustrated in a meaningful manner.

The SOM algorithm automatically places a set of reference vectors—also called model vectors—into the input data space so that the data set is approximated by the model vectors. The model vectors are constrained to a (usually two-dimensional) regular grid that in effect forms an "elastic network": after application of the SOM algorithm, the map follows the data in a nonlinear fashion. The SOM algorithm obtains simultaneously *a clustering* of the data onto the model vectors and a *nonlinear projection* of the input data from the high-dimensional input space onto the two-dimensional, ordered lattice formed by the model vectors. Furthermore, the mapping defined by the model vectors can later be used to project new, unseen data onto the clusters or the map display.

3 Data and features

3.1 Selected verbs and the corpora

For this study, we selected 25 frequent Finnish verbs, shown in Table 1. All the occurrences of these verbs, a total of about 500,000 samples, were extracted from two sub-corpora of written standard Finnish (CSC, 2001), namely 'Newspapers' (13.6 million words) and 'Books' (4.1 million words). The morphological analysis of the data was conducted by the Conexor FDG, a functional dependency parser for Finnish (Tapanainen and Järvinen, 1997).[2] No effort was made to detect possible parsing errors. Furthermore, the verbs were not disambiguated semantically.

3.2 Contextual morphological features

To describe a verb we collected features from a window of a single word to the right of the target word, regardless of the syntactic function of the word in question. The following morphosyntactic properties were included: 1) a set of case endings (Table 2), 2) adpositions, i.e., postpositions and prepositions (see example 2), 3) adverbs, and 4) two nominal forms of verbs (i.e., the 1st and the 3rd infinitive, see example 3). Thus, the only information the system 'knows' is whether the next word immediately to the right of the target word bears some of the selected features.

One criterion for selecting the feature set was to prefer overtly marked features that can be identified by a system able to carry out some segmentation. These include case endings and the endings of the nominal forms of verbs. Moreover, by including 'adverb' and 'adposition' as features we assumed that the system is already capable of distinguishing parts of speech.

In Finnish, the primary means for coding various functional and syntactic dependencies between the verb and the nominal constituents in a clause is the case-marking system. The case endings are added to stems, as shown in (1), where the subject is marked by the nominative, and the end point of the movement indicated by the verb *ajaa* 'to drive' is marked by the illative.

(1)	Lapse-t	ajoivat	kaupunki-in.
	child	drive	city
	N-PL-NOM		N-SG-ILL

'The children drove to the city.'

In addition to the so-called local cases, location or movement can be coded by using an adposition (2), too. Adpositions include postpositions (PSP) and prepositions (PRE). They take a complement either in genitive or in partitive.

(2)	Lapse-t	ajoivat	kaupunki-a	kohti.
	child	drive	city	toward
			N-SG-PTV	PSP

'The children were driving towards the city.'

Adpositions, adverbs, and local cases are grammatical categories that are at least partly func-

[2] ©Conexor oy and Anu Airola, <http://www.conexor.fi>

Verb	freq.	Translation
1. PSYCHOLOGICAL STATES AND PROCESSES, MODAL VERBS		
alkaa	17,987	begin, start, commence
haluta	14,637	want; wish; like (how do you like your tea); care for (would you care for a cup of tea?)
nähdä	14,746	see, spot
pitää	38,299	I) hold, keep, retain II) must, have to, be compelled to, be obliged to, shall; be supposed to
saada	57,815	get; receive; may, can, be allowed to, be permitted to; have to
tietää	12,678	know, be aware of, be conscious of
voida	46,474	can, be able, be capable, may
2. MOTION VERBS		
ajaa	6,372	drive; ride; run
käydä	14,447	go; walk
laskea 1	5,150	I) tr. lower, let down, haul down II) itr 1. fall, decline, go down, drop, decrease, diminish
laskea 2		count; calculate
lähteä	12,921	go, leave
mennä	18,486	go
nousta	10,889	rise; go up; ascend; climb; arise
päästä	14,809	a) get (into/out of); arrive, reach b) be let (into a room); be allowed (to enter); be admitted
tulla	43,972	come; arrive; become, get, turn, grow into
tuoda	6,840	bring
3. SPEECH ACT VERBS		
kertoa	21,474	tell; narrate, relate, recount; inform
puhua	10,389	speak, talk; tell; say; discuss
sanoa	43,015	say
vastata	7,554	answer, respond, reply, give an answer, make a reply, counter
4. ACTION VERBS		
käyttää	12,035	use (a knife)
tehdä	27,195	do; perform; commit; make
5. VERBS OF POSSESSION		
antaa	14,269	give; let; allow
kuulua	12,425	belong; be heard, be audible
ottaa	18,501	take

Table 1: The verbs divided into five categories according to Pajunen's ontological-semantic verb classification (Pajunen, 1999). The number after the verb shows the frequency of occurrences of the verb in the studied corpora. Translations are from (Hurme et al., 1984). Note that many of the verbs are polysemous and different senses should, in fact, be placed into different categories.

tionally overlapping. However, as examples (1) and (2) show, they construe the situation differently. The two clauses show an aspectual opposition; the illative in (1) indicates that the action is completed, while the postposition *kohti* 'towards' in (2) expresses duration without specifying completion.

The feature set used also includes two nominal (non-finite) verb forms, namely the 1st and the 3rd infinitive. The infinitives function in a sentence as nouns (3).

(3)	Halua-n	lähte-ä	syömä-än.
	want	go	eat
	V-PRES-SG1	V-INF1	V-INF3-ILL
	'I want to go to eat.'		

In Finnish, the lower-level constituents, e.g., NPs (noun phrases), are predominantly head-final, the order being modifier-before-head. Adjectives agree in number and case with the head-noun when they occur as attributes (4). Due to this redundancy, the function of a dependent NP can be inferred before hearing or seeing the head of the phrase.

(4)	Emmi	ajaa	punaise-lla	auto-lla.
	Emmi	drive	red	car
	N-PROP	V-SG3	A-SG-ADE	N-SG-ADE
	'Emmi drives in a red car.'			

In the clause-level, the basic, or default, word order is SVO. According to (Hakulinen et al., 1980), subjects precede finite verbs in 61% of all sentences in standard written prose. However, as Vilkuna (1989) points out, clause-level word order in Finnish shows great freedom; for example, in a simple sentence consisting of a subject, an object, a verb and one or two adverbials, all permutations are at least grammatically possible.

4 Experiments

4.1 Experiment 1: A map of 25 Finnish verbs

The purpose of this experiment was to obtain an overview of the verbs selected for study. In particular, we were interested in whether the contextual window of only one word and the set of morphological features would contain sufficient information for obtaining automatically a functional ordering or clustering for the verb base forms. As a basis for comparison, we used Pajunen's verb classification (see Table 1).

4.1.1 Data and preprocessing.

Let V be the set of 25 verbs, $v_i \in V$, $i = 1 \ldots 25$; and M the set of 18 morphological tags, $m_j \in M, j = 1..18$. In the data vector $\mathbf{x}_i$ for verb v_i a separate input dimension is reserved for representing each morphological tag. The value of $x_{i,j}$ is the frequency of context $v_i m_j$ in the corpus normalized by the frequency of verb v_i. In effect, $\mathbf{x}_i$ contains estimates of the conditional probabilities $Pr(m_j(t+1)|v_i(t))$ for each j (t is an index over the sequence of words). In other words, the vector $\mathbf{x}_i$ describes how typical each of the morphological tags is in the word following the verb v_i.

4.1.2 The organization and clustering of the data.

After forming the data vectors, various data analysis methods can be applied. We organized the verb representations automatically with the SOM algorithm, using the SOM ToolBox (Vesanto et al., 2000) for the data analysis tasks and for the visualizations. Default values were used for SOM learning parameters. K-means[3] was utilized for clustering the map vectors after creation of the map.

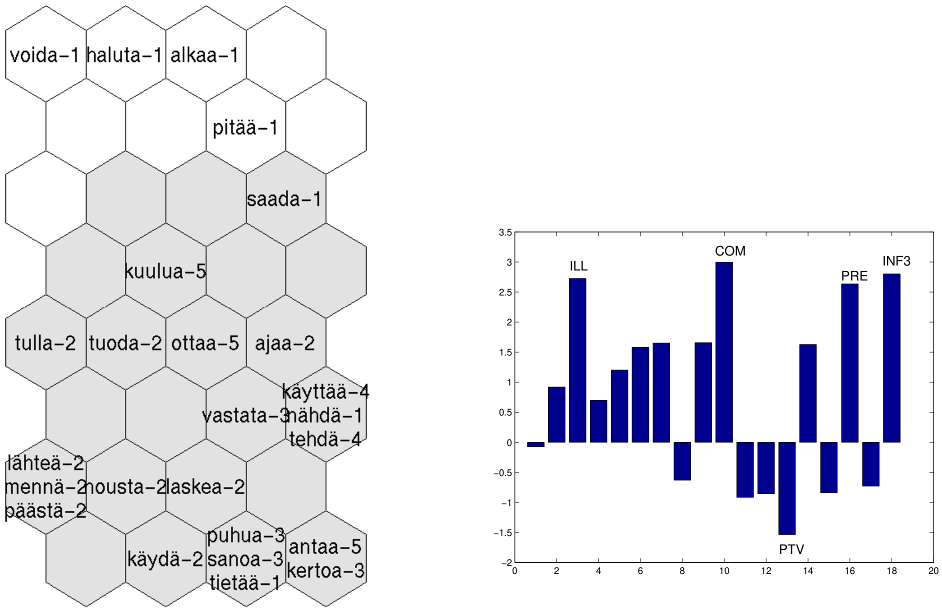
4.1.3 Results.

For each verb, the best-matching map unit (bmu) was found by locating the model vector closest to the verb vector. The verb was then written onto the bmu on the map lattice in Fig. 1. The data vectors of verbs related to two selected map units are visualized in Fig. 2. The values of individual features have been plotted on the map in Fig. 3.

When comparing the obtained functional organization to Pajunen's ontological-semantic verb classification, there seems to be a reasonable amount of agreement: verbs with the same category number are rather often found near each other on the map display.

However, even the perceived differences can be meaningfully justified. For example, the common aspect of *puhua-sanoa-tietää* ('to speak,' 'to say,' 'to know') is that they are cognitive verbs related to communicating or possessing in-

[3]K-means algorithm is obtained as a special case of the SOM by disregarding the neighborhood in the organization.



(a) (b)

Figure 1: Analysis of the 25 verbs. (a) The SOM lattice where each hexagon marks a map unit. The map units belonging to cluster 2 are shaded (bottom of the map). The best-matching unit, bmu, has been automatically searched for each of the 25 verb vectors, and each verb has been written on its bmu. For example, the bmu for the verbs *lähteä, mennä,* and *päästä* was the unit near the lower left corner of the map. The number after the verb refers to the verb class of Table 1. (b) The difference between the map unit of verbs *lähteä, mennä, päästä* and the rest of the map for each input feature. The values are normalized with standard deviation of the feature. If the value for a feature (say COM) is above zero, that feature is more frequent in this map unit than in general, and below zero means that it is here less frequent than in general. The features that most distinguish this unit from other units are ILL, COM, PRE, INF3, and the lack of PTV.

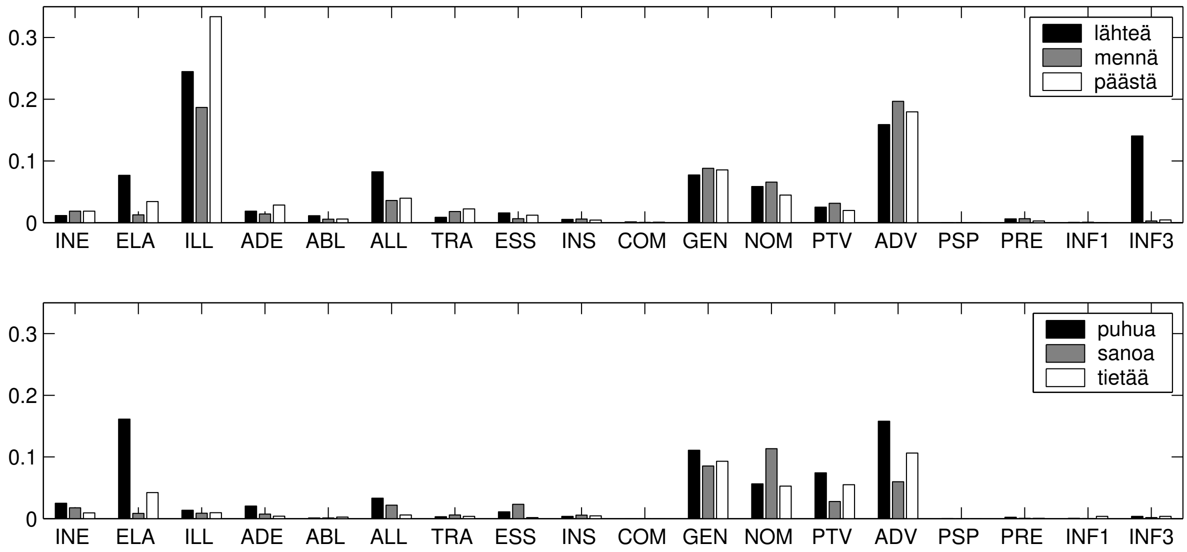


Figure 2: Data vectors for verbs in two different map units: the value for each morphological form is visualized as the height of the bar.

Case		Endings	Meaning	Example
		GRAMMATICAL CASES		
nominative	NOM	- (pl. -t)	basic form	auto ’car’
partitive	PTV	-a/-ä; -ta/-tä;	indefinite	maito+a ’(some) milk’
		-tta/-ttä	quantity	vet+tä ’(some) water’
genitive	GEN	-n; -den, -tten	possession	auto+n ’of the car’
		LOCAL CASES		
inessive	INE	-ssa/-ssä	inside	auto+ssa ’in the car’
elative	ELA	-sta/-stä	out of	auto+sta ’out of the car’
illative	ILL	-Vn,	into	auto+on ’into the car’
		-hVn,		maa+han ’into the country’
		-seen, -siin		Lontoo+seen ’to London’
adessive	ADE	-lla/llä	on; instrument	pöydä+llä ’on the table’
ablative	ABL	-lta/ltä	off	pöydä+ltä ’off the table’
allative	ALL	-lle	onto	pöydä+lle ’onto the table’
		OTHER CASES		
essive	ESS	-na/-nä	state	opettaja+na ’as a teacher’
translative	TRA	-ksi	change of state	opettaja+ksi ’(become) a teacher’
comitative	COM	-ine-	accompanying	vaimo+ine+ni ’with my wife’
instructive	INS	-n	(idiomatic)	jala/n ’on foot’

Table 2: The cases used as input features (cf. (Karlsson, 1987): 22–23).

formation (CONTENT). In the neighboring unit, *antaa-kertoa* (’to give,’ ’to tell’) describe concrete or abstract change of possession (RECIPIENT). Moreover, *käyttää-nähdä-tehdä* (’to use,’ ’to see,’ ’to do’) describe actions towards an object (PATIENT).

Pajunen’s classification exhibits a certain ontological view of the world. As the comparison with our experiment suggests, an alternative perspective emerges when one concentrates on functional features and on what is typical in actual language use.

4.2 Experiment 2: A map of 4764 verb forms

The purpose of the second experiment was to study more closely the 21 verbs in the second cluster (the lower part of the map obtained in the first experiment, see Fig. 1). In particular, of interest were the various inflected verb forms, since it is to be expected that different forms of the same verb may have quite different behavior profiles.

The data vectors were formed similarly as in experiment 1, the only difference being that a vector was now collected per each verb form. To avoid unreliable estimates verb forms with fewer than 20 occurrences were discarded. As a result,

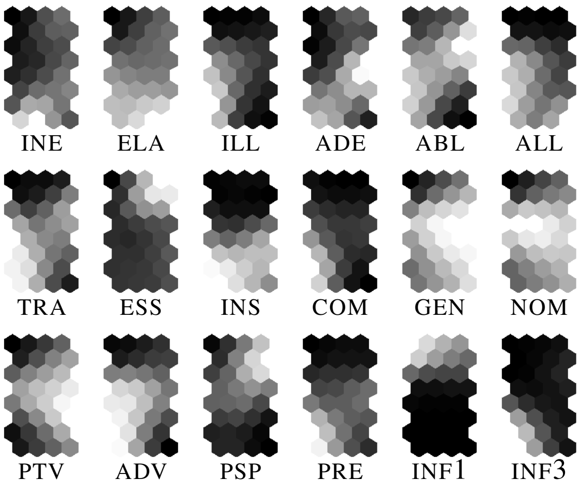


Figure 3: The values of individual features have been visualized on the map—bright colors signal large values. For example, the most prominent area of the feature INF3 is the lower left corner of the map. By comparing the visualizations of two features, dependencies between them can be identified. For example, features COM, PRE, and INF3 seem strongly correlated.

4764 data vectors were obtained. Next, a map of 11×14 units was automatically created with the SOM ToolBox. Best-matching units were then located for each data vector.

4.2.1 Results

To obtain an overall view of the map organization of the verb form map, we plotted all the occurrences of verb forms of a single base form onto a single map display. Several examples are depicted in Fig. 4. Fig. 5 shows more closely the behavior of different verb forms, such as the contrast between finite and non-finite forms of verbs.

The spread of a verb shows how coherently the forms of that verb behave in terms of the applied feature selection. By comparing spreads of two different verbs, say *mennä* and *lähteä*, one quickly obtains some overview of the differences and similarities between the verbs—in this case, that the verbs are rather similar. However, a difference is implied since the verbs fall on adjoining areas instead of overlapping.

By looking at the values of the morphological features on the map (not shown here) it can be seen that the feature INF3 (3rd infinitive) is especially strong on the area overlapping with the verb *lähteä* but not on the adjoining area of *mennä*. In fact, the illative form of the 3rd infinitive can be used after all motion verbs, indicating an action which is about to begin, as in (5):

	Lähden / Menen	nukku-ma-an.
(5)		V-INF3-ILL
	'I'll go to sleep.'	

However, as the experiment shows, the 3rd infinitive illative is particularly typical of the verb *lähteä*. In contrast, the verbs *mennä* and *päästä* are used more often to indicate a 'pure' change of location.

5 Conclusions and Discussion

On the outset it may have seemed surprising that we decided to use only the word immediately following the verb for collecting information. After all, (1) many verbs take several arguments, and (2) since the word order in Finnish is free—or better, discourse-conditioned (Vilkuna, 1989)—the arguments may in some cases appear quite far apart from the verb, either before or after the verb.

Nevertheless, in Experiment 1, comparison between the map ordering and the pre-existing verb categorization shows that the selected morphosyntactic features of the next word indeed are statistically quite indicative of the functional-semantic properties of the verb, *if the information is accumulated over a large number of individual samples*. In such an accumulation, the typical cases will dominate and the less typical cases, many of which may be random (noise), are less frequent and become averaged out over time.

One might think that studying larger contexts would give a rather different view of the verbs. In fact, we repeated Experiment 1 so that both the preceding and succeeding words surrounding the verb were considered (using separate feature dimensions), with very similar results. It is possible that due to the unconstrained word order, wider contexts, if utilized in a similar manner, provide only a marginal amount of additional information. Nevertheless, especially when clustering rare words a wider context might become useful and worthwhile. Also additional features might be worth considering.

With respect to computational language modeling, the obtained clustering seems promising. In the future, larger numbers of verbs, including rare verbs, should be examined, and the approach should be extended to other kinds of words. Furthermore, the obtained clustering should be quantitatively evaluated in a specific task.

The visualizations obtained with the SOM can be extremely valuable, e.g., for a linguist interested in *actual language use*. This kind of method allows tackling larger problem areas than before. As seen from the simple examples presented here, the possibility to obtain an overview of the behavior of a group of words with a glance at the data may draw attention to some interesting differences that are otherwise hard to detect, and perhaps give rise to new perspectives or hypotheses regarding the linguistic phenomena.

References

Branimir Boguraev and James Pustejovsky, editors. 1996. *Corpus Processing for Lexical Acquisition*. The MIT Press, Cambridge, Massachusetts.

Eugene Charniak. 1993. *Statistical Language Learning*. MIT Press.

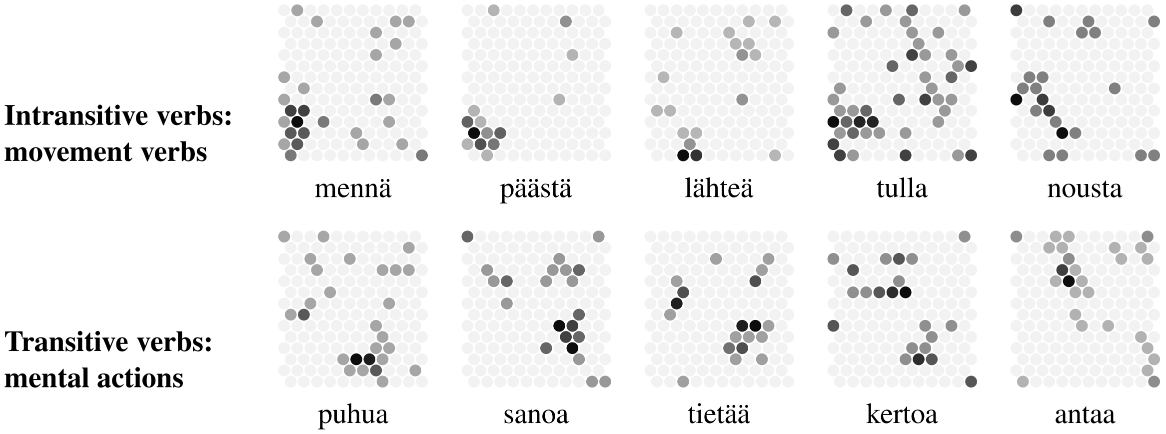


Figure 4: Verb distributions on the verb form map. The first plot depicts the occurrences of verb forms with the base form *mennä*—dark colour corresponds to a large number of occurrences. Each plot can be viewed as a "higher-order fingerprint" of the verb—similar fingerprints correspond to similar distributions of word forms in terms of their morphological contexts. The plots on the first row depict verbs that were found on the lower left side of the map in the previous experiment (see Fig. 1), whereas the second row consists of the verbs in the lower right corner of the same map.

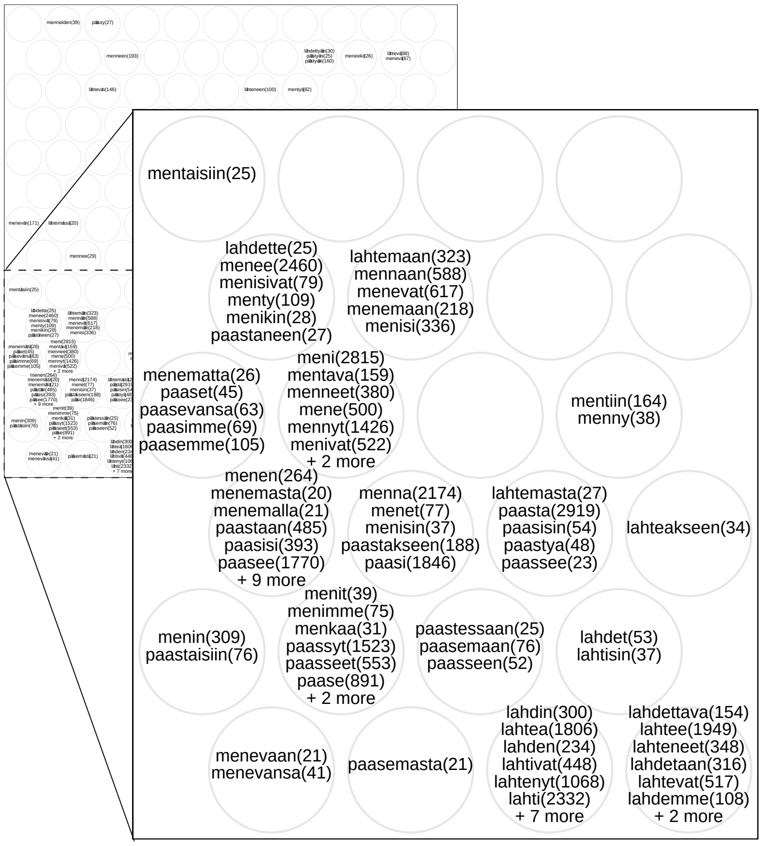


Figure 5: A close-up on the lower left corner of the map. The individual verb forms whose base form was *lähteä* 'to leave,' *mennä* 'to go,' or *päästä* 'to get into/out of' have been plotted on the map to their best-matching units (the scandinavian characters *ä* and *ö* are shown here as *a* and *o*). The numbers of occurrences of each form are shown in parenthesis. The verb forms in this area are mostly finite forms of the verbs. They seem to be in contrast to the non-finite forms, which appear as another, weaker cluster in the upper right corner of the map (not visible).

William Croft. 1999. Typology and cognitive linguistics. In Theo Janssen and Gisela Redeker, editors, *Cognitive Linguistics: Foundations, Scope, and Methodology*, pages 61–94. Mouton de Gruyter, Berlin.

2001. CSC—tieteellinen laskenta oy. Finnish Language Text Bank: Corpora 'Books' and 'Newspapers'. http://www.csc.fi/kielipankki/.

S. Finch and N. Chater. 1992. Unsupervised methods for finding linguistic categories. In I. Aleksander and J. Taylor, editors, *Artificial Neural Networks, 2*, pages II–1365–1368. North-Holland.

Adele Goldberg. 1995. *Constructions. A Construction Grammar Approach to Argument Structure*. Cognitive Theory of Language and Culture. The University of Chicago Press, Chicago.

Auli Hakulinen, Fred Karlsson, and Maria Vilkuna. 1980. *Suomen tekstilauseiden piirteitä: kvantitatiivinen tutkimus*. Publications of the Department of General Linguistics, Yliopistopaino, Number 6, Helsinki.

Simon Haykin. 1999. *Neural Networks—A Comprehensive Foundation*. Prentice Hall, 2nd edition.

Timo Honkela, Ville Pulkki, and Teuvo Kohonen. 1995. Contextual relations of words in Grimm tales analyzed by self-organizing map. In F. Fogelman-Soulié and P. Gallinari, editors, *Proceedings of ICANN-95, International Conference on Artificial Neural Networks*, volume II, pages 3–7, Paris. EC2 et Cie.

Timo Honkela. 1997. *Self-Organizing Maps in Natural Language Processing*. Ph.D. thesis, Helsinki University of Technology, Espoo, Finland.

Raija Hurme, Riitta-Leena Malin, and Olli Syväoja. 1984. *Uusi suomi-englanti suursanakirja. Finnish-English General Dictionary*. WSOY, Porvoo.

Fred Karlsson. 1983. *Suomen kielen äänne- ja muotorakenne*. WSOY, Porvoo. See also http://www.ling.helsinki.fi/~fkarlsso/bookpaprep.html.

Fred Karlsson. 1987. *Finnish Grammar*. WSOY, Juva, the second edition edition.

Samuel Kaski, Timo Honkela, Krista Lagus, and Teuvo Kohonen. 1998. WEBSOM—self-organizing maps of document collections. *Neurocomputing*, 21:101–117.

Teuvo Kohonen, Jussi Hynninen, Jari Kangas, and Jorma Laaksonen. 1996. SOM_PAK: The Self-Organizing Map program package. Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science, January.

Teuvo Kohonen. 1982. Self-organizing formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69.

Teuvo Kohonen. 1995. *Self-Organizing Maps*. Springer, Berlin.

Caludia Leacock, Geoffrey Towell, and Ellen M. Voorhees. 1996. Towards building contextual representations of word senses using statistical models. In Branimir, Boguraev, and James Pustejovsky, editors, *Corpus Processing for Lexical Acquisition*, pages 97–113. The MIT Press, Cambridge, Massachusetts.

Christopher D. Manning and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, Massachusetts.

George A. Miller and Walter G. Charles. 1991. Contextual correlates of semantic similarity. *Language and Cognitive Processes*, 6(1):1–28.

Anneli Pajunen. 1999. *Suomen verbirektiosta. Verbin argumenttirakenteen jäsenten valinta*. Yleisen kielitieteen julkaisuja 1. Åbo Akademis tryckeri, Åbo.

H. Ritter and T. Kohonen. 1989. Self-organizing semantic maps. *Biological Cybernetics*, 61:241–254.

John Sinclair. 1991. *Corpus, Concordance, Collocation*. Oxford University Press, Oxford.

Michael Stubbs. 1996. *Text and Corpus Analysis: Computer Assisted Studies of Language and Culture*. Blackwell, Oxford.

Pasi Tapanainen and Timo Järvinen. 1997. A non-projective dependency parser. In *Proceedings of the 5th Conference on Applied Natural Language Processing*, Washington, D.C., April. Association for Computational Linguistics.

Michael Tomasello, editor. 1998. *The New Psychology of Language*. Lawrence Erlbaum Associates, Mahwah, NJ.

Juha Vesanto, Johan Himberg, Esa Alhoniemi, and Juha Parhankangas. 2000. Som toolbox for matlab 5. Report A57, Helsinki University of Technology, Neural Networks Research Centre, Espoo, Finland, April.

Maria Vilkuna. 1989. *Free Word Order in Finnish. Its Syntax and discourse functions*. Suomalaisen kirjallisuuden seura, Helsinki.

Jakub Zavrel and Jorn Veenstra. 1995. The language environment and syntactic word-class aquisition. In *Proceedings of the Groningen Assemly on Language Aquisition (GALA95)*, pages 365–374.