

**Antti Arppe**

**Frequency Considerations in Morphology, Revisited  
– Finnish Verbs Differ, too**

**Abstract**

In the mid-1980s Fred Karlsson published two papers in Eastern European linguistic journals, which have remained notoriously difficult to get hold of, but for the persistent researcher contain the seeds of a very interesting hypothesis worth further pursuit. Karlsson's argument in these papers was that different noun lexemes in Finnish prefer different inflectional forms on the basis of their basic meanings. Accordingly, they also prefer different individual morphological features associated with these forms (Karlsson 1985, 1986). Karlsson also argues that these preferences, designated as focal forms, would not be lexeme-specific, but would rather exhibit groupings according to some, rather general semantic classifications. In contrast, Karlsson suggests that Finnish verbs would not exhibit similar preferential differentiation based on their semantic groupings, and that they would in fact behave as one group, all sharing the same, single focal inflected form. The purpose of this paper is to suggest that Karlsson was probably both right and wrong, i.e. Finnish verbs do differ like nouns according to their semantic groupings, when their inflectional profiles are considered in their entirety with multivariate statistical techniques. However, Karlsson's original stance may still hold when taken literally, i.e. Finnish verbs apparently do not exhibit other than idiosyncratic individual focal forms.

**1. Introduction and background**

As Finnish is a morphologically rich language, its nouns, verbs, and adjectives can theoretically have thousands of different (inflected) word forms. Generally given figures are just over 1,850 for a noun, just below 6,000 for an adjective, and approximately 20,000 for a verb. Though such inflectional paradigms can be viewed as both full and their individual members as equal according to some earlier views (by H. Seiler, P. H. Matthews, A. Carstairs and M. Aronoff, as reviewed by Karlsson 1985: 137), Karlsson (1985: 151) has argued on the contrary that a paradigm should rather be viewed as a gradient continuum. Thus, instead of a two-way division into forms that can and do exist and forms that cannot exist,

*A Man of Measure  
Festschrift in Honour of Fred Karlsson, pp. 175–189*

i.e. as either grammatical or ungrammatical, one should see the paradigm as a profile representing the relative frequencies of its members, reflecting if not varying degrees of grammaticality then varying acceptability or naturalness. The relative frequencies are determined as a function of (morpho-)syntactic and semantic properties of the root lexeme (Karlsson 1985: 149–150; 1986: 23), via the respective semantic and syntactic roles they are used in. In such a profile, the most frequently used forms, which Karlsson coins as *focal forms*, (1986: 23) are at one end, and the fully deficient forms, i.e. *lacunae*, at the other, with the remaining forms falling in between these two extremes. Nevertheless, the dichotomous view of members of a paradigm as either existing and grammatical or not, without gradiency according to frequency considerations, appears to have persisted, see e.g. Stump (1998: 1–2, 16–17) and Carstairs-McCarthy (1998: 324–325).

Karlsson (1986: 20–23) suggests that these profiles need not be fully different and idiosyncratic for each individual lexeme. Instead, there are instead some general types, which are determined by the “basic” meanings of the root lexemes, such as HUMAN AGENTS, LOCATIONS, MASS nouns, and TEMPORAL nouns. As Arppe (2001) has observed in the case of two such general semantic groupings of Finnish nouns, namely the MASS and TEMPORAL nouns, they are not fully uniform with respect to the preferred morphological features but exhibit several subgroupings, suggesting that Karlsson’s principle could be seen to apply at several levels of semantic classification and grouping. Furthermore, according to Arppe, some inflected forms and the associated morphological features would be common and characteristic for the entire semantic grouping, differentiating the group as a whole from others at the same level of classification. On the contrary, some other inflected forms and morphological features would be differentiating ones, specific to the semantic subgroupings.

In contrast to nouns, Karlsson has argued that Finnish verbs would not exhibit such clustering around meaning-related focal forms as nouns do, but would rather share one single focal point, namely the ACTIVE INDICATIVE THIRD PERSON SINGULAR form, roughly corresponding to English forms such as ‘He/she/it is/does/sees/thinks/says/feels/wants/...’ (Karlsson 1986: 25–26). Karlsson furthermore judges whatever differences there may be as individual idiosyncratic forms rather than as representative of any semantic grouping. Karlsson based his argument on corpus-based observations of a set of 12 Finnish verbs that are all very frequent and are typically also highly polysemous. As a consequence it may very well be

that the different morphological profiles corresponding to the multiple different senses blur the picture. Thus, without word-sense disambiguation these lexemes will appear to converge around this least marked, lowest common denominator of a focal form, even more so when observed in newspaper texts with an emphasis on the reporting narrative text type. Furthermore, Karlsson looked at individual inflected forms instead of the entire inflectional profiles or their component features. This approach might not sufficiently expose combinatorial association effects that can rather be discovered with multivariate statistical techniques.

Approaching this issue from the bottom up, Arppe (2002) focused on two synonymous verbs deliberately selected for having as similar syntactic valency structures, semantic meaning potentials and relative frequencies as possible. The two verbs selected, *mieltiä* and *pohtia* ‘think, reflect, ponder’, belong to the group of THINK verbs. Arppe discovered that such near-synonymous verbs can clearly have a number of purely inflectional differences. In fact, very similar observations can already be found in the detailed description of the Finnish modal verb system by Kangasniemi (1992).

In a further study, Arppe (2005) worked upwards and expanded this set of THINK verbs to the nearest, more general level in the semantic classificational hierarchy, namely to the COGNITION verbs. Contrary to Karlsson's judgement, the conclusions of this study support the argument that Finnish verbs, as well as nouns, do have characteristic inflectional profiles which becomes evident when studied in their entirety using a multivariate statistical method, instead of looking at one morphological feature or form at a time. Furthermore, these inflectional profiles are not arbitrary but correspond in a regular manner with semantic classifications. Moreover, this regularity was observed to apply at several levels of granularity at the less general, bottom end of the underlying semantic classification hierarchy, with semantically rather closely related groups of verbs. However, the results do not give a clear indication of whether lexemes belonging to different semantic groups of verbs at some level of classification would indeed have in their inflectional profiles one or more particular focal forms. This is what Karlsson can be interpreted to have strictly meant, i.e. which (full) word form(s) would be characteristic to the entire semantic group and not merely amount to individual idiosyncrasies. Nevertheless, it still remains open whether the more general semantic groupings of verbs towards the top of the hierarchy would exhibit similar differentiation.

The purposes of this paper are two-fold: 1) to fill in the gap in previous studies and demonstrate that the general principle proposed by Karlsson of an association between preferred inflectional forms and meaning also applies for the top-most, most general semantic groupings of verbs, and thus 2) to demonstrate that this principle applies in a similar fashion to all different levels of granularity of semantic classification and grouping of verbs from the top down to individual synonym groups. Such results would also establish that morphological features, at least in the case of Finnish, are a contextual category in its own right in differentiating semantically similar words.

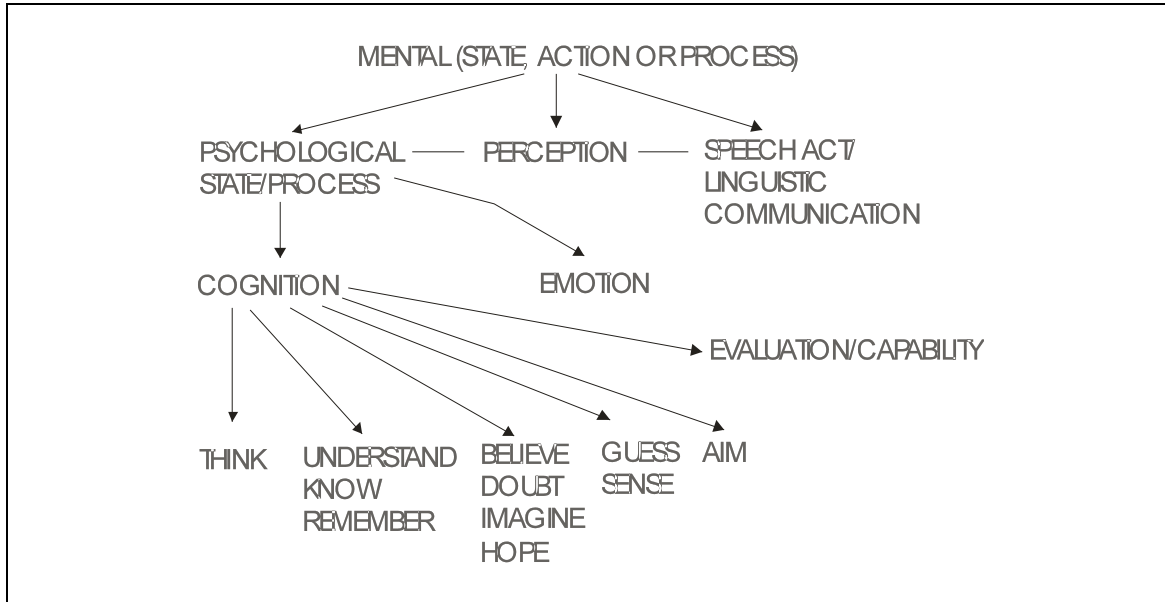
## 2. Corpora and methods

In lack of a Finnish WordNet, the most comprehensive available classification of Finnish verbs is the hierarchy presented by Pajunen (2001: 56–57, 63), a modified version of Pajunen (1982). Since Pajunen provides prototypical examples rather than comprehensive listings for the bottom-level subgroupings in either of her classifications, the authoritative *Suomen kielen perussanakirja*, i.e. the ‘Basic Dictionary of Finnish’, known in its electronic version as *CD-perussanakirja* (Haarala & al. 1997), has also been extensively consulted. The existing relevant work concerning Finnish verbs and their morphological preferences, mainly undertaken by Arppe (but see also Vanhatalo 2005 for adjectives and verbs and Jantunen 2001, 2004 for adjectives), has hitherto focused on synonym groups like the THINK and UNDERSTAND verbs,<sup>1</sup> and the nearest more general group, namely the COGNITION verbs. It would be natural to continue upwards to the top-most general levels in the semantic classificational hierarchy, namely to the closely related PSYCHOLOGICAL STATE/PROCESS, PERCEPTION and VERBAL OR NON-VERBAL LINGUISTIC COMMUNICATION (SPEECH ACT) verbs. Together these form the group of MENTAL STATE, ACT OR PROCESS verbs. As VOLITION verbs formed a distinct subgroup in Pajunen’s earlier classification (1982: 147–206, 334–336), they are treated as such in this study, too. Furthermore, the PERCEPTION verbs were in Pajunen’s earlier

---

<sup>1</sup> In this paper, synonym groups will be denoted by their most frequent and/or prototypical member, e.g. the THINK verbs, whereas all the more general groupings will be denoted by a descriptive concept, e.g. the COGNITION verbs. The names given for these concepts are the author’s translations of the original Finnish terms by Pajunen (1982, 2001).

classification designated as a subgroup of the PSYCHOLOGICAL verbs rather than as a top-level group of their own. Accordingly, PSYCHOLOGICAL verbs are viewed as a specific group in the second study in this paper.



**Figure 1.** The semantic classification hierarchy of MENTAL verbs according to Pajunen (2001)

Previous experience has shown that the visualization of the results of the correspondence analysis with more than 30–40 items becomes in practice illegible. Thus, both studies were limited to the 25–30 most frequent lexemes, the exact number varying to allow for an even inclusion of all semantic subgroupings. Table 1 provides the list of verbs included in the first study in this paper, which focuses on the comparison of the top-level PSYCHOLOGICAL, PERCEPTION or VERBAL OR NON-VERBAL LINGUISTIC COMMUNICATION (SPEECH ACT) groups, indicating also their specific semantic subgroupings. Table 2 provides the less frequent verbs that were included in the second study, in addition to those in Table 1. As some of the selected verbs are polysemous or homonymous, so that they may belong to several different semantic groups, such parallel senses are also indicated in the Tables.

The research corpus included texts from Helsingin Sanomat (1995), the largest daily newspaper of Finland, amounting to about 1.7 million words. Only body text was included in the corpus, excluding headers, captions, and so forth, which are sources of significant repetition in the relatively short newspaper articles. The corpus was automatically morpho-

syntactically analyzed with an implementation of the Functional Dependency Grammar parser (Tapanainen & Järvinen 1997) for Finnish (FI-FDG), developed by Connexor <www.connexor.com>. In order to cope with analyzing the multitude of inflected forms in Finnish, the FI-FDG parser provides a set of morphological features, which can be directly used as such in the subsequent stages of the analysis. After the automatic analysis, no manual disambiguation or correction of the results was undertaken, and the remaining morpho-syntactic ambiguity (affecting slightly less than 10% of the observed instances of the studied verbs) was excluded from the ensuing statistical analysis. This decision may naturally have some influence on the results.

Verbs and their English translations ( <u>X</u> ~ suffix indicating semantic classification)	Semantic subgroup	Semantic supergroup	Polysemous/homonymous senses	Frequency
<i>sanoa_V</i> 'say'	Verbal (communication)	(Linguistic) Communication	–	3053
<i>Kertoa_V</i> 'tell'	Verbal	Communication	'multiply'	2221
<i>Haluta_W</i> 'want'	Volition	Psychological (state or process)	–	1393
<i>Kuulua_P</i> 'be heard'	Perception	Perception	'belong to'	1088
<i>Esittää_V</i> 'propose'	Verbal	Communication	'show, play'	1049
<i>Nähdä_P</i> 'see'	Perception	Perception	–	949
<i>Laskea_P</i> 'count'	Perception	Perception	'lower/go down'	749
<i>Vaatia_V</i> 'demand'	Verbal	Speech act	–	732
<i>uskoa_C</i> 'believe'	Cognition	Perception	–	718
<i>tietää_C</i> 'know'	Cognition	Perception	–	676
<i>ilmoittaa_V</i> 'announce'	Verbal	Communication	–	659
<i>Puhua_V</i> 'speak'	Verbal	Communication	–	649
<i>vastata_V</i> 'answer'	Verbal	Communication	'correspond to'	649
<i>Katsoa_P</i> 'look'	Perception	Perception	–	640
<i>Yrittää_W</i> 'try'	Volition	Psychological	–	636
<i>Todeta_V</i> 'state'	Verbal	Communication	–	634
<i>Tuntea_E</i> 'feel'	Emotion	Psychological	–	579
<i>hyväksyä_E</i> 'accept'	Emotion	Psychological	–	573
<i>Nimittää_V</i> 'nominate'	Verbal	Communication	–	545
<i>näyttää_P</i> 'appear'	Perception	Perception	'show'	521
<i>kirjoittaa_V</i> 'write'	Verbal	Communication	–	517
<i>myöntää_V</i> 'concede'	Verbal	Communication	–	515
<i>aikoa_W</i> 'aim'	Volition	Communication	–	514
<i>Tarjota_V</i> 'offer'	Verbal	Communication	–	512
<i>näkyä_P</i> 'be seen/apparent'	Perception	Perception	–	503

**Table 1.** List of MENTAL verbs, i.e. PSYCHOLOGICAL, PERCEPTION and LINGUISTIC COMMUNICATION (VERBAL or NON-VERBAL) verbs, included in this study, together with their subgroupings and frequencies

Verbs and their English translations (_X ~ suffix indicating semantic classification)	Semantic subgroup	Polysemous/ homonymous senses	Frequency
<i>löytää</i> P 'find, discover'	Perception	–	487
<i>pyrkii</i> W 'aim at'	Volition	–	453
<i>tuntua</i> P 'feel'	Perception	–	435
<i>toivoa</i> W 'wish'	Volition	–	356
<i>tutkia</i> P 'study, explore'	Perception	–	355
<i>arvioida</i> P 'assess, judge'	Perception	–	315
<i>suunnitella</i> W 'plan, aim to'	Volition	–	305
<i>ymmärtää</i> C 'understand'	Cognition	–	301
<i>kuulla</i> P 'hear'	Perception	–	265
<i>ajatella</i> C 'think'	Cognition	'aim'	253
<i>Muistaa</i> C 'remember'	Cognition	–	235
<i>pohdii</i> C 'ponder'	Cognition	–	235
<i>arvella</i> C 'guess, surmise'	Cognition	–	235
<i>epäillä</i> C 'suspect, doubt'	Cognition	–	234
<i>huomata</i> P 'recognize, realize'	Perception	–	226
<i>pelätä</i> E 'fear'	Emotion	–	223
<i>kokea</i> E 'feel, experience'	Emotion	–	211
<i>vastustaa</i> W 'oppose, object to'	Volition	–	209

**Table 2.** List of PSYCHOLOGICAL STATE/PROCESS and PERCEPTION verbs included in the further study, together with their frequencies and indication of their semantic subgroupings.

In the first, study, altogether, 21,274 instances of the 25 selected verbs were observed in the research corpus, ranging from 3,053 occurrences of the most frequent *sanoa* 'say', down to 503 occurrences of *näkyä* 'be seen/apparent'. In the second study of the 31 selected verbs in the PSYCHOLOGICAL and PERCEPTION groups, 14,872 instances were observed in the corpus, ranging from 1,393 for the most common *haluta* 'want', down to 209 for *vastustaa* 'oppose, object to'. One must note that as the corpus was not semantically analyzed in order to discern polysemous or homonymous cases of the studied verbs, some of this semantic ambiguity will most probably have an effect on the results, especially in a case where the different senses essentially differ in some respect, e.g. *näyttää* as both the intransitive 'appear' and the transitive 'show', or *vastata* as both the transitive 'answer' and the intransitive 'correspond to'.

Correspondence Analysis (Lebart & al. 1998), an established statistical method especially popular in French corpus-based linguistics, was then used for the analysis and visualization of the similarities and differences of the inflectional profiles within the studied verb groupings. A peculiar characteristic of the correspondence analysis is that it visually organizes the data in the categories into central and peripheral instances. The increasing distance of any representative of either category from the

origin corresponds to a higher degree of differentiation as compared with the other members with respect to their co-occurrences with the data in the other category. In this way it differs from Self-Organizing Maps and Cluster Analysis, which group items but establish neither a specific center nor a periphery.

### 3. Results

The results of the correspondence analyses of the first and second studies are presented in Figures 2 and 3, respectively. Approximate English translations for the Finnish lexemes in the figures have been given in Tables 1 and 2 above. In the statistical analysis, such individual morphological features were excluded which had either 1) less than 50 observed instances overall in the research corpus or 2) the maximum observed frequency of such a morphological feature together with any individual of the studied lexemes was less than 10.

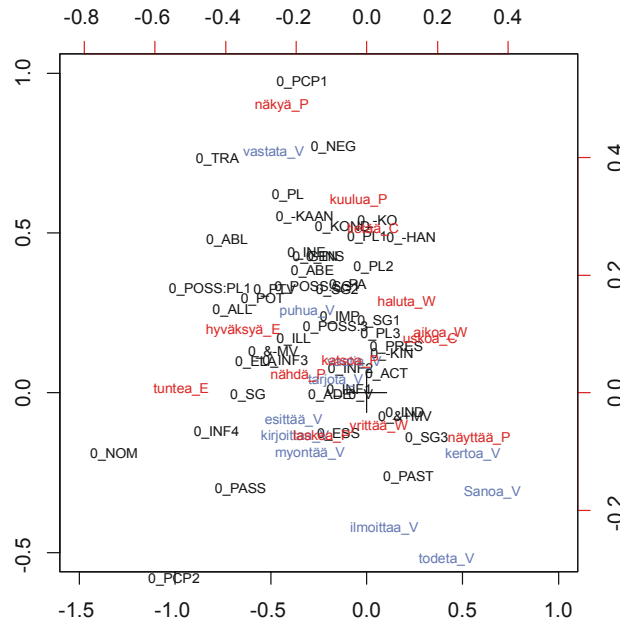
#### LEGEND for MORPHO-SYNTACTIC TAGS in Figures 2 and 3:

Node-specific feature	0_XXX
Finite form	&+MV
Non-finite form	&-MV
Infinitive forms 1-5	INF1, INF2, INF3, INF4, INF5
Present and Past Participle forms	PCP1, PCP2
Active and Passive voice	ACT, PASS
Negated form	NEG
Indicative, Conditional, Potential and Imperative mood	IND, KOND, POT, IMP
Present and Past tense	PRES, PAST
Person and Number of finite forms	SG1, SG2, SG3, PL1, PL3, PL3
Cases (for nominal-like infinitives and participles)	NOM, GEN, PTV, ...
Singular and Plural number (for nominal-like forms)	SG, PL
Possessive suffixes (used in non-finite clauses):	POSS:SG1-PL3
Clitics	-KO (questions), -KIN ('also'), -KAAN ('neither')

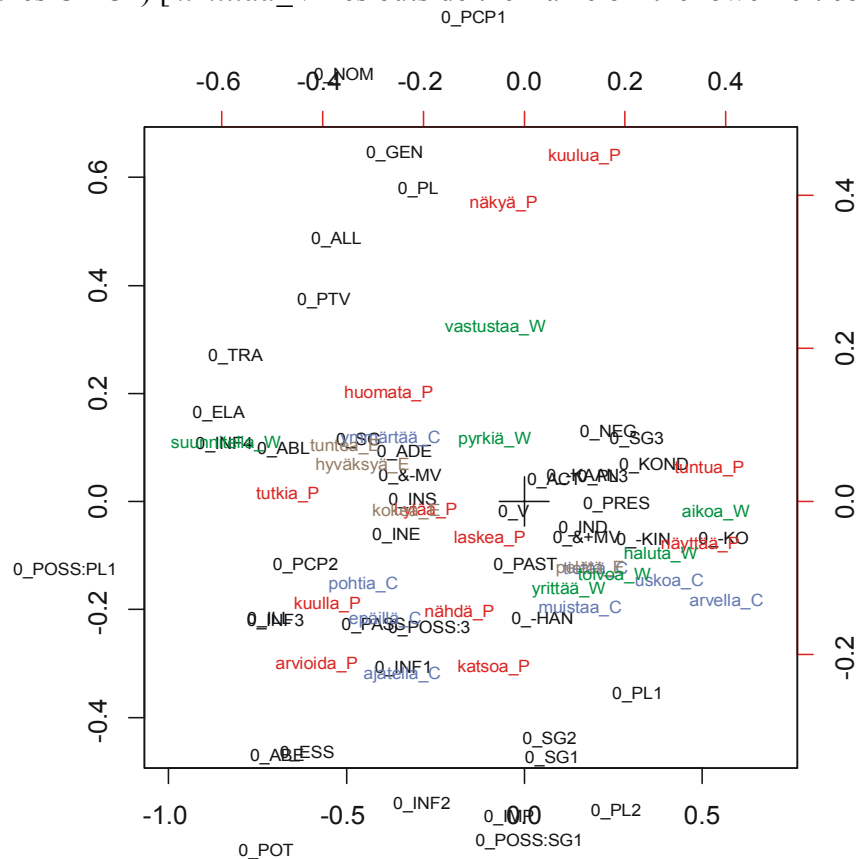
#### LEGEND for SEMANTIC CLASSIFICATIONS in Figures 2 and 3 plus color codes in Figure 3 (when relevant):

COGNITION lexeme (blue color)	xxx_C
EMOTION lexeme (brown color)	xxx_E
VOLITION lexeme (green color)	xxx_W
PERCEPTION lexeme (red color)	xxx_P
VERBAL COMMUNICATION (SPEECH ACT) lexeme	xxx_V





**Figure 2.** Correspondence analysis of the PSYCHOLOGICAL PROCESS/STATE and PERCEPTION (red color) and VERBAL AND NON-VERBAL COMMUNICATION (blue color) lexemes (axes C1-C2) [*nimittää\_V* lies outside the frame off the lower left corner]



**Figure 3.** Correspondence analysis of the studied PSYCHOLOGICAL STATE/PROCESS and PERCEPTION verbs (axes C1-C2)

#### 4. Discussion

As noted in earlier similar research (Arppe 2005) regarding the interpretation of correspondence analysis, one must be careful about making statements concerning the association of individual morphological features – and especially their combinations – with individual studied lexemes. Based on a more detailed study of the original corpus data, it may turn out that some of the most outlying morphological features or their combinations, which would appear to be strong candidates as distinguishing traits, have actually only a few instances in the corpus. For instance, in Figure 3 the FIRST PERSON PLURAL POSSESSIVE suffix lying to the far left has in the research corpus overall only 3 occurrences, associated one each with *suunnitella* ‘plan’, *tuntea* ‘know/recognize’ and *katsoa* ‘watch’. On the other hand the PRESENT PARTICIPLE has 1,116 occurrences, spread over all but one of the studied PSYCHOLOGICAL and PERCEPTION verbs.

Likewise, out of several morphological features lying in the same direction only some may co-occur to a significant degree among the studied lexemes, and only some together with the lexeme lying in the same direction. However, in the case of the verbs studied in this paper, which all belong to the most frequent cases in their respective groups, the visually apparent associations seem to have a relatively solid basis in terms of actual frequency in the research corpus. For instance, of the visual co-occurrences in Figure 2 the PRESENT PARTICIPLE co-occurs in the corpus with the TRANSITIVE case 38 times, of which 21 with *vastata* ‘answer’, and the PAST PARTICIPLE, PASSIVE voice and NOMINATIVE case co-occur 397 times, of which 204 with the outlying *nimittää* ‘nominate/appoint’, both arising from the idiosyncratic construction *[X on] nimitetty [Y:stä] vastaavaksi* ‘[X has been] appointed [as the person] answering, i.e. responsible [for Y]’ common in business and political news. Moreover, of the visual co-occurrences in Figure 3 the PRESENT PARTICIPLE and the GENITIVE case co-occur in the corpus 181 times with 27 out of 31 of the studied COGNITION verbs, of which a third with *kuulua* ‘be seen/belong’.

In general, the morphological features are best interpreted as forming a background profile by which they together as a group organize and relate the studied lexemes with respect to each other, in a manner which will be demonstrated below to suggest a few clear semantic interpretations. It is therefore the positioning and grouping of the studied lexemes which is of the foremost linguistic interest and significance in the diagrams.

Nevertheless, there are some exceptions of a significant association of an individual morphological feature together with some particular lexemes, which will be indicated in the interpretations below.

What is most significant in Figure 2 is that the MENTAL STATE, ACT OR PROCESS verbs seem to form two groups along the vertical axis, with the “extrovert” LINGUISTIC COMMUNICATION verbs<sup>2</sup> more towards the top and the “introvert” PSYCHOLOGICAL STATE/PROCESS and PERCEPTION verbs<sup>3</sup> more towards the bottom. In fact, a t-test on the distributions of the vertical positions in the diagram indicates that the two groupings are disparate to a statistically significant degree.<sup>4</sup> This would strongly indicate that the two semantic groupings differ in regard to the morphological forms and features in which they are used, i.e. their inflectional profiles. However, the farthest outlying PRESENT and PAST PARTICIPLE forms have been shown above to be associated somewhat with individual lexemes, i.e. *vastata* ‘answer’ and *nimittää* ‘nominate’, respectively. On the other hand, it appears both on the basis of the visual results and the frequencies in the research corpus that NEGATION and the FIRST and SECOND PERSON SINGULAR and PLURAL features are somewhat more associated with the PSYCHOLOGICAL verbs, whereas the PAST tense and THIRD PERSON SINGULAR are somewhat more associated with the LINGUISTIC COMMUNICATION verbs. Nevertheless, the singular person features are rather in the core than in the periphery, which may reflect their dual central role in current newspaper genre, with the reporting of events in the third person and direct quotations in the first person. Of all the observed features, the FIRST INFINITIVE is the most central, positioned right in the origin, which would challenge Karlsson’s original assertion regarding the focal form for Finnish verbs, even more so as it is also the traditional citation form of verbs in Finnish dictionaries.

In Figure 3 it seems that the representatives of the different subgroups of PSYCHOLOGICAL STATE/PROCESS and PERCEPTION verbs do not form any visually distinct groupings but are instead quite interleaved, with perhaps the COGNITION, EMOTION and VOLITION verbs converging tighter around the center than the PERCEPTION verbs. This is understandable as the verbs in question, in addition to the prototypical traits by which the groups have been arranged, also share peripheral semantic traits across the subgroup boundaries. Viewed along these peripheral traits the positioning of the

<sup>2</sup> Designated as MENTAL-PROCESS verbs in Pajunen’s earlier classification (1981).

<sup>3</sup> Grouped together as MENTAL verbs in Pajunen’s earlier classification (1981).

<sup>4</sup> Two-tailed Student’s t-test with a p-value = 0.035.

various PSYCHOLOGICAL verbs does in fact start to make sense, with the horizontal axis representing 1) the degree of determination and purposefulness of psychological state or process itself or its result, decreasing from left to right, and the vertical axis representing 2) the degree of agentivity of the PSYCHOLOGICAL process, decreasing from top to down.

Thus, along the vertical axis, one can see a transition from an intransitive process with no obligatory sentient recipient or observer, such as *kuulua* ‘be heard’ and *näkyä* ‘be seen’, through transitive psychological processes with a sentient but “passive” experiencer, such as *huomata* ‘notice’, *tuntea* ‘know/recognize’, *tuntua* ‘feel’, *kuulla* ‘hear’ and *nähdä* ‘see’, ending up with consciously initiated and maintained psychological processes with “active” agents, such as *arvioida* ‘assess’, *ajatella* ‘think’ and *katsoa* ‘watch’. The only lexeme not really fitting neatly in this scheme is *vastustaa* ‘oppose’. Along the horizontal axis, we can see a shift from determined psychological effort with a determined purpose, such as *suunnitella* ‘plan’, *tutkia* ‘study/observe’ and *arvioida* ‘assess’, through intentionality such as *pyrkiä*, *yrittää* ‘try’ and *vastustaa* ‘oppose’, to a vague or undetermined, or even doubtful or wishful attitude, such as *tuntua* ‘feel’, *aikoa* ‘aim’, *näyttää* ‘appear’, *arvella* ‘guess’ *haluta*, ‘want/wish’ and *uskoa* ‘believe’.

Regarding the morphological features, the interpretation of the vertical axis in terms of agentivity works well, as the FIRST and SECOND PERSON SINGULAR and PLURAL, the FIRST PERSON POSSESSIVE SUFFIX (which represents the agent in non-finite constructions), and the loosely associated IMPERATIVE mood features are at the very bottom in the most agentive end of the diagram. On the other hand, the PRESENT PARTICIPLE, along with the various associated cases, is at the least agentive end at the top in the diagram, which fits with the use of this feature in non-finite attributive constructions, e.g. *näkyvä piirre* ‘apparent/observable feature’. The THIRD PERSON SINGULAR could be deemed as a shared and undifferentiating feature, being close to the center, whereas the FIRST INFINITIVE is now off center, this time supporting Karlsson’s original assertion on the focal form of Finnish verbs, in contrast to Figure 2.

Finally, with regards to the reliability of the results, as the lexemes studied in this paper belong to most frequent group of Finnish verbs, many of them are polysemous to the extent that it has an apparent influence on the observed morphological behavior of a part of the lexemes and thus on the full validity of the above observations. This is evident especially in the

idiosyncratic cases, which have also been observed often as quite frequent, and where the particular usage can strictly speaking be understood as belonging to the studied semantic groups through a meaning-chain. Take for instance. *vastaava* ‘responsible [person]’, morphologically constructed as the PRESENT PARTICIPLE of *vastata*, meaning literally ‘answer somebody regarding a question’, but by extension also ‘answer to somebody concerning a responsibility’<sup>5</sup>. The frequent use of *vastaava* in the latter sense of ‘(be) responsible for’ would account for its positioning in Figure 2 among the intransitive PERCEPTION verbs, as would the use of *näyttää* in the sense of ‘show’ rather than ‘appear’ explain its positioning in the same Figure among the transitive VERBAL COMMUNICATION verbs. Therefore, as the studied verbs in the research corpus were not analysed and classified as to their different senses, the conclusions presented in this paper would certainly benefit from validation and refining with the help of a semantically disambiguated corpus.

## 5. Conclusions

In conclusion, the interpretations of the results presented above firstly confirm the argument, contrary to Karlsson's judgement, that Finnish verbs, as well as nouns, do have characteristic inflectional profiles, which becomes evident when studied in their entirety using a multivariate statistical method such as correspondence analysis, instead of looking at one morphological feature or form at a time. Furthermore, as has been observed in earlier studies these inflectional profiles are not arbitrary but correspond in a regular manner with semantic hierarchical classifications or other relevant semantic dimensions. Secondly, in combination with earlier studies and the results presented in the paper, this regularity has been observed to apply at in practice all levels of granularity of the underlying semantic classification of verbs, from the closest level of similarity, i.e. synonymy (Arppe 2005), through increasingly more general groupings all the way to the top-most levels of the semantic hierarchy studied in this paper. However, as in earlier studies the results do not give a clear indication of whether lexemes belonging to different semantic groups at some level of classification would indeed have in their inflectional profiles one or more particular focal forms, as Karlsson can be interpreted to have

---

<sup>5</sup> The meaning potential of the verbs in question being apparently quite similar in both Finnish and English.

strictly meant, which word form(s) would be characteristic to the entire semantic group and not merely amount to individual idiosyncrasies that have been observed above. Finally, the results would also warrant for the (re-)consideration of the status of members (cells) of a paradigm along a gradient rather than dichotomous basis, as was originally suggested by Karlsson.

## References

- Arppe, Antti (2001). Focal points in frequency profiles – how some word forms in a paradigm are more significant than others in Finnish. In *Proceedings of the 6th Conference on Computational Lexicography and Corpus Research (COMPLEX-2001)*, June 28-30, 2001, pp. 1–8. Birmingham: University of Birmingham.
- (2002). The usage patterns and selectional preferences of synonyms in a morphologically rich language. In Annie Morin & Pascale Sébillot (eds) *JADT-2002. 6th International Conference on Textual Data Statistical Analysis, March 13-15, 2002*, Volume 1, pp. 21–32. Rennes: INRIA.
- (2005). The role of morphological features in distinguishing semantically similar words in Finnish – A study of cognitive verbs. In *Proceedings from the Corpus Linguistics Conference Series*, Volume 1, no. 1. Available on-line at URL: <http://www.corpus.bham.ac.uk/PCLC/>.
- Carstairs-McCarthy, Andrew (1998). Paradigmatic Structure: Inflectional Paradigms and Morphological Classes. In Spencer & Zwicky (eds.), pp. 322–334.
- Haarala, Risto & Marja Lehtinen (eds.) (1997) *CD-Perussanakirja*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 94. Helsinki: Edita.
- Jantunen, Jarmo Harri (2001) Tärkeä seikka ja keskeinen kysymys: mitä korpuslingvistinen analyysi paljastaa lähisynonyyeistä? *Virittäjä* 105.2: 170–192.
- (2004). *Synonymia ja käännessuomi: korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännskielen leksikaalisiin erityispiirteisiin*. University of Joensuu Publications in the Humanities 35. Joensuu: University of Joensuu.
- Kangasniemi, Heikki (1992) *Modal Expressions in Finnish*. Studia Fennica, Linguistica 2. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Karlsson, Fred (1985) Paradigms and word forms. *Studia gramatyczne* VII: 135–154.
- (1986) Frequency considerations in morphology. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 39.1: 19–28.
- Lebart, Ludovic, André Salem & Lisette Berry (1998) *Exploring Textual Data*. Text, Speech and Language Technology 4. Dordrecht: Kluwer Academic Publishers.
- Pajunen, Anneli (1982). *Suomen kielen verbien leksikaalinen kuvaus*. Licentiate Thesis. Department of Finnish and Linguistics, University of Turku.
- (2001). *Argumenttirakenne: Asiaintilojen luokitus ja verbien käyttäytyminen suomen kielessä*. Suomi 187. Helsinki: Suomalaisen Kirjallisuuden Seura.

- Spencer, Andrew & Arnold M. Zwicky (eds.) (1998) *The Handbook of Morphology*. Blackwell Handbooks in Linguistics. Cambridge, MA: Blackwell Publishers.
- Stump, Gregory T. (1998). Inflection. In Spencer & Zwicky (eds.), pp. 13–43.
- Vanhatalo, Ulla (2005). *Kyselytestit synonymian selvittämisessä – Sanastotietoutta kielenpuhujilta sähköiseen sanakirjaan*. PhD Dissertation. Helsinki: Department of Fenno-Ugric Studies, University of Helsinki. Also available on-line from URL: <http://ethesis.helsinki.fi/julkaisut/hum/suoma/vk/vanhatalo/kyselyte.pdf>

## Corpora

Helsingin Sanomat 1995. ~2 million words of Finnish newspaper articles published in January 1995. Compiled by the Research Institute for the Languages of Finland [KOTUS] and CSC – Center for Scientific Computing, Finland. Available on-line at URL: <http://www.csc.fi/kielipankki/>.

Contact information:

Antti Arppe  
Department of General Linguistics  
University of Helsinki  
P.O. Box 9 (Siltavuorenpenger 20 A)  
FI-00014 University of Helsinki  
antti(dot)arppe(at)helsinki(dot)fi  
<http://www.ling.helsinki.fi/~aarppe>