

Ulla Vanhatalo

Is there Synonymy between Finnish Idioms—and How to Describe or Measure it?¹

Abstract

Synonymy is generally understood as referring to similarities of sense between single lexical items. The potential synonymy of idioms² has been mostly ignored in lexical semantics. The synonymy of idioms is, however, implicated in e.g. the occasional practice of cross-referencing in Finnish idiom dictionaries. This paper sets out to see whether the testing of a native-speaker population using a questionnaire method could disclose novel aspects of the synonymy of idioms in Finnish. The present findings are compatible with the idea that synonymy is found between idioms, which in turn raises some tentative options for improving synonymy-related glossing in idiom dictionaries. In addition, the project also yielded a large, open-access database of synonymous Finnish idioms.

¹ I wish to thank Elke Gehweiler, Tarja Riitta Heinonen, Marja Nenonen, Oksana Petrova, Jarno Raukko and Maria Vilkkuna for their support. The preliminary results were presented at *Kielitieteen päivät 2006* (Vanhatalo 2006), many thanks for the useful comments from the public during this event. Special thanks belong to Anne Takkunen, who organized the experiment for the students of Jämsän Kr. Kansanopisto (Jämsä Christian Folk High School) in March 2006. Finally, I want to sincerely thank the informants who voluntarily invested their time and energy in this experiment.

² Idiom in the sense of a conventional, noncompositional, syntactically and/or lexically frozen lexical item longer than one word (e.g. Nenonen 2002: 8); this should be understood as a definition of a prototypical idiom, with a substantial variety of exceptions. Recently there have also been attempts to make idiom concepts more flexible (e.g. Penttilä 2006).

1. Introduction

1.1 Synonymy—a property restricted to single lexemes?

Despite the extensive literature on the properties of idioms in a large number of languages, very little attention has been paid to the possible existence of synonymy between phrasal lexical items. The overall emphasis of the current literature suggests an implicit assumption that synonymy is restricted to single words (e.g. Chrystal 2003).

Recent studies in Slavic, German and French (Danell 1997) linguistics have addressed the issue of synonymy in phraseology. Most Russian studies have been based on either structural distinctions or on dividing phrases into full and partial synonyms (Čerkasova 1991). Some studies have also followed Kunin's (1996) conception by dividing phraseological synonyms into groups with respect to their differences in meaning, connotation, and style, forming groups named ideographic synonyms, stylistic synonyms and stylistic-ideographic synonyms (Soshnikova 2006). There are at least two synonymy dictionaries on Russian phraseology (Žukov 1987; Birih et al. 1998/2001). Many German studies have taken advantage of the ongoing large-scale electronic lexicography (especially the collocations project in the German language, Akademie der Wissenschaften), which has made it possible to examine extensive German corpora in order to disclose the contextual conditions that lead to fixed and dynamic semantic convergence or divergence of idioms (Hümmer 2004). Studies of this kind have repeatedly emphasized the need for much more fine-grained lexical information to be obtained between semantically closely related lexical units in Natural Language Processing (Hümmer 2004; see also Edmonds 1999; Edmonds and Hirst 2002).

In Finnish studies the topic of synonymous idioms has been peripherally touched upon in studies on euphemisms and word taboos (Nirvi 1944, Varis 1998; see also Rapola 1944; Tuomola 1935). There are also references to synonymy between phrasal lexical items in Finnish-German idiom studies (e.g. Korhonen 1995 or Hyvärinen 1996). The idea of synonymy between phrasal lexical items has not been specifically targeted in any of these studies.

While the synonymy between idioms may be intuitively obvious and theoretically interesting as such, its practical significance becomes clear in

the lexicographical context. Even in the absence of systematic analytic lexicographic data, there has been occasional cross-referencing within Finnish phraseological and idiom dictionaries (see e.g. Virkkunen 1974; Kari 1993; Korhonen 2001). Synonymy is also recognized in dictionaries between single headwords and their glosses which often are phrases.

The present study was motivated by the apparent lack of theoretical and pragmatic data in facing this lexicographic challenge. The study aims at proving with empirical populations testing, that implicit and intersubjective³ understanding of synonymy between idioms exists, as it does for single word lexical items. The study draws attention to synonymy as a phenomenon, and proposes ways to make current glossing practice in dictionaries more explicit and systematic.

1.2 A preliminary data collection on synonymy in the field of Finnish idioms

A preliminary data collection was made in winter 2005–2006⁴. The objective was to find idiom pairs or groups that could be used for further studies, as well as for cross-referencing in future Finnish idiom dictionaries.

Most of the database was collected from the idiom lists in Marja Nenonen's Ph. D. thesis (2002) and Erkki Kari's idiom dictionary (1993). Additional data was collected from many other available sources of idioms (e.g. Virkkunen 1974, Rekiaro 1998). Since there are as yet no corpus tools that could identify idioms and/or their synonymy in Finnish, the data was repeatedly scanned by the analysts reading the idiom lists and dictionaries.

³ For intersubjectivity, see e.g. Raukko 1999.

⁴ When the preliminary results from the data collection process were presented at *Kielitieteen päivät 2006*, there were signs of interest especially from Joensuu idiom researchers towards co-operative and open-access data collection. The currently used database can be found at www.helsinki.fi/people/ulla.vanhatalo. The database is open and free to use for anyone interested. At the moment, there are about 1370 idioms forming over 300 synonymic pairs or groups in the database.

2. The research setting

2.1 Method

This study employed an experimental questionnaire method⁵ which is based on an intersubjective view of linguistic understanding, i.e., lexical knowledge resides in the minds of native speakers. The genuine characteristics of lexemes may hence be best exposed by strategies that attempt to elicit the tacit (i.e. hidden) knowledge of the native speakers. The present method utilizes tailored questions related to the target lexicon to make this implicit knowledge explicit and potentially measurable.

A method of this kind has been successfully used in several recent studies on lexical semantics, including experiments on synonymy (e.g. Vanhatalo 2005), polysemy (e.g. Raukko 1999), acknowledgements (e.g. Colston 2002), and idioms (e.g. Nenonen 2002). Similar strategies have been occasionally used in linguistic research for more than half a century (Nirvi 1944; see also Rapola 1944).

2.2 Material and participants

A subset of the database was selected with a pilot test.⁶ The final test set included 25 idioms that comprised 11 synonymous pairs and one three-idiom group⁷ (see Appendix 1). The questionnaire also included some ‘fillers’, idioms with strikingly different meanings. Odd additions of this kind are thought to single out informants who are not attentive enough. The informant group consisted of 94 native speakers of Finnish, whose age range was 16–25 yrs. (average 16.5 yrs.), and of whom 53,6 % were female. Most participants were born and raised in Southern, Western or Central Finland. All participants answered the same questions, which took 6–24 minutes. For practical details of the testing, see e.g. Vanhatalo 2005.

⁵ For more about the method and the practical test organization see Vanhatalo 2005. Some of the test settings in this study were inspired by Marja Nenonen 2002.

⁶ The pilot testing took place in the graduate seminar of the Department of General Linguistics at Helsinki University in January 2006.

⁷ Some other idioms which are not presented here were also studied at the same time.

2.3 The questions and their results

2.3.1 The first section

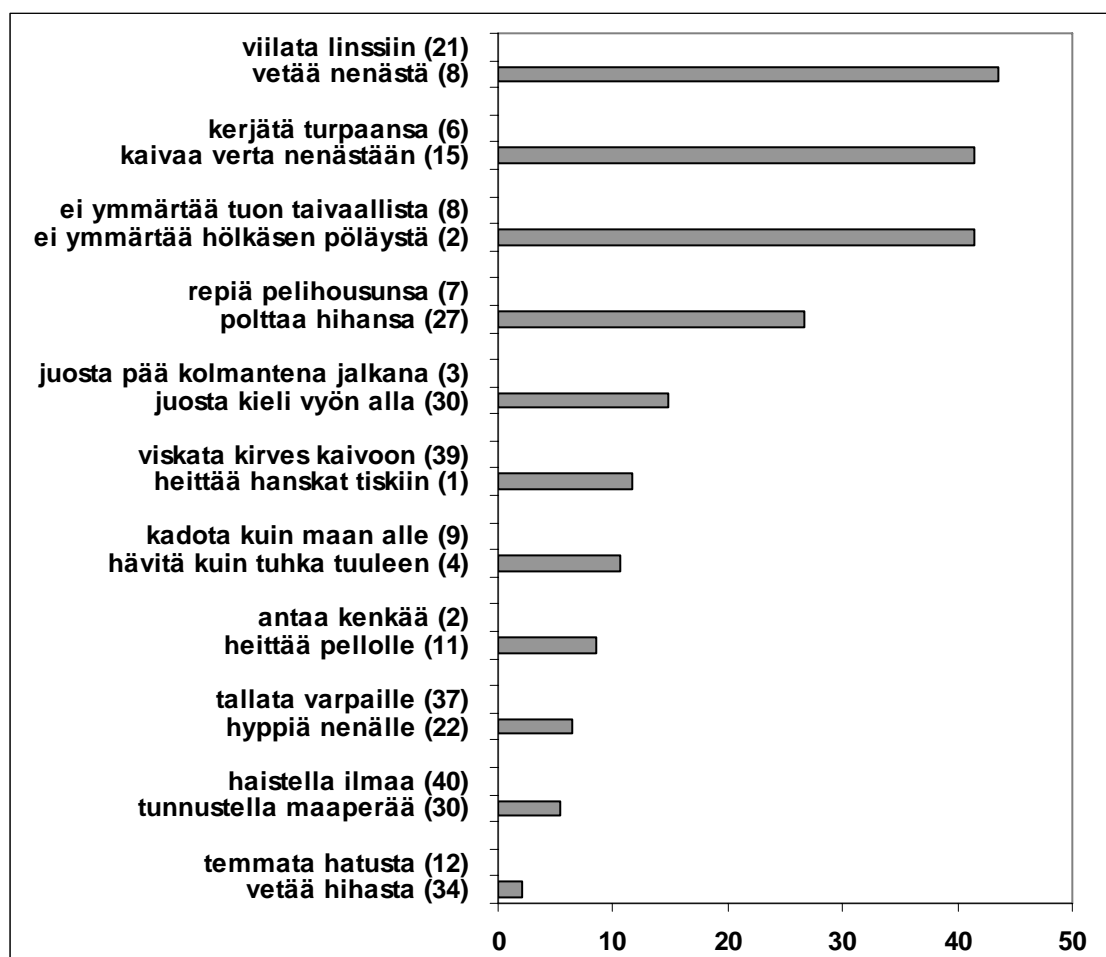


Table 1. Summary of the answers to the first test section showing the percentage of informants who used exactly similar phrasing for explaining both idioms in each synonymic idiom pairs. Note that the synonymic pairs were not placed consecutively, so the informant is presumably generating phrasings/descriptions anew for each idiom, rather than simply copying the phrasings from the preceding idiom. The numbers in parentheses after each idiom show how many informants gave an answer to the given idiom (i.e. empty and “I do not know” answers excluded).

The first test section (11 idiom pairs) included open-ended questions which aimed to test whether similarities between idioms could be revealed by

having the informant freely explain the meanings of the idioms. Surprisingly, a large number of the 94 students chose identical (25–40 students) or near-identical (37–52 students) phrases for both idioms (Table 1). For instance, the idiom pair *repiä pelihousunsa* ‘lit. rip one’s game pants; flare up’ and *polttaa hihansa* ‘lit. burn one’s sleeves; flare up,’ were glossed as *menettää hermonsa* or *mennä hermot* ‘loose one’s nerves;’ *raivostua* ‘get mad;’ *suuttua* ‘get angry’ and *hermostua* ‘get nervous’. This is strongly suggestive of the hypothesis that genuine synonymy exists between idioms. The other end of Table 1 reveals another interesting feature: The meaning of idioms may change over time. Older speakers normally presume that the idioms *temmata hatusta* ‘pull from a hat; guess’ and *vetää hihasta* ‘pull from a sleeve; guess’ denote the same action (‘to guess’). It turned out from the students’ results that the informant group conceived the meaning of the idiom *vetää hihasta* rather often as ‘to stop someone to ask for help or advice’, which is an emerging, contemporary meaning of this idiom. Finally, a considerable proportion of the idioms were not familiar to the informants.

2.3.2 The second section

In the second test section the informants rated the synonymy between idiom pairs on a scale from 0 to 5. The responses from all informants are strikingly similar, giving ratings between 3 and 5 (Table 2). These results hence further corroborate the idea that synonymy is a fundamental and distinct property of idioms.

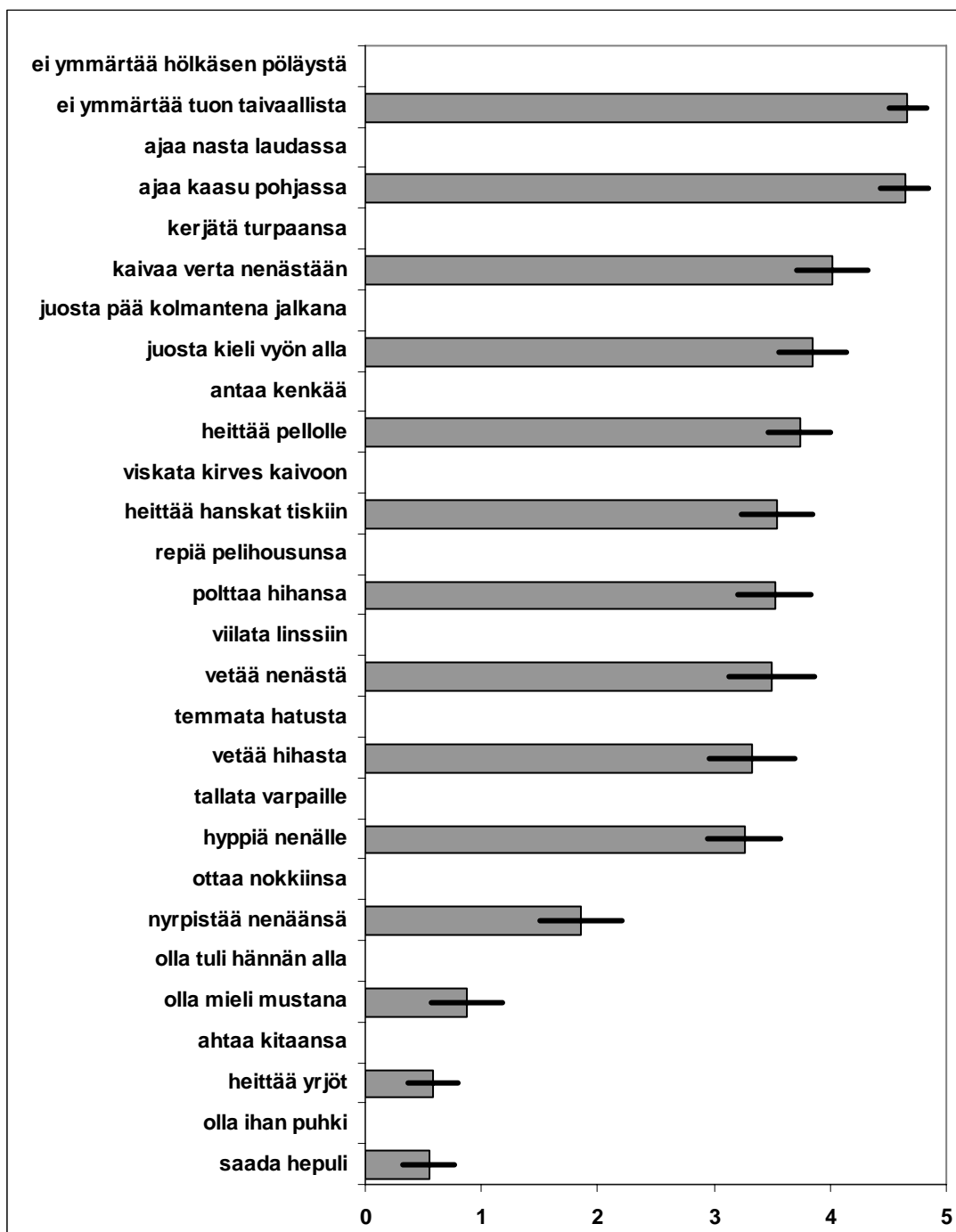


Table 2. Summary of the answers to the second test section showing the level of similarity between each idiom pair. Error bars depict 95 percent confidence intervals in each answer. The lowest three pairs are clearly recognized as fillers with a similarity index below 1, while the semifiller is rated below 2.

2.3.3 The third section

The third test section was an open-ended test where the informants were requested to explain the differences in the meaning or use of single idioms. Here two idioms were contrasted, while the first section sought the similarity between idioms. The idioms selected here were all common Finnish idioms used to denote the act of driving a car very fast.

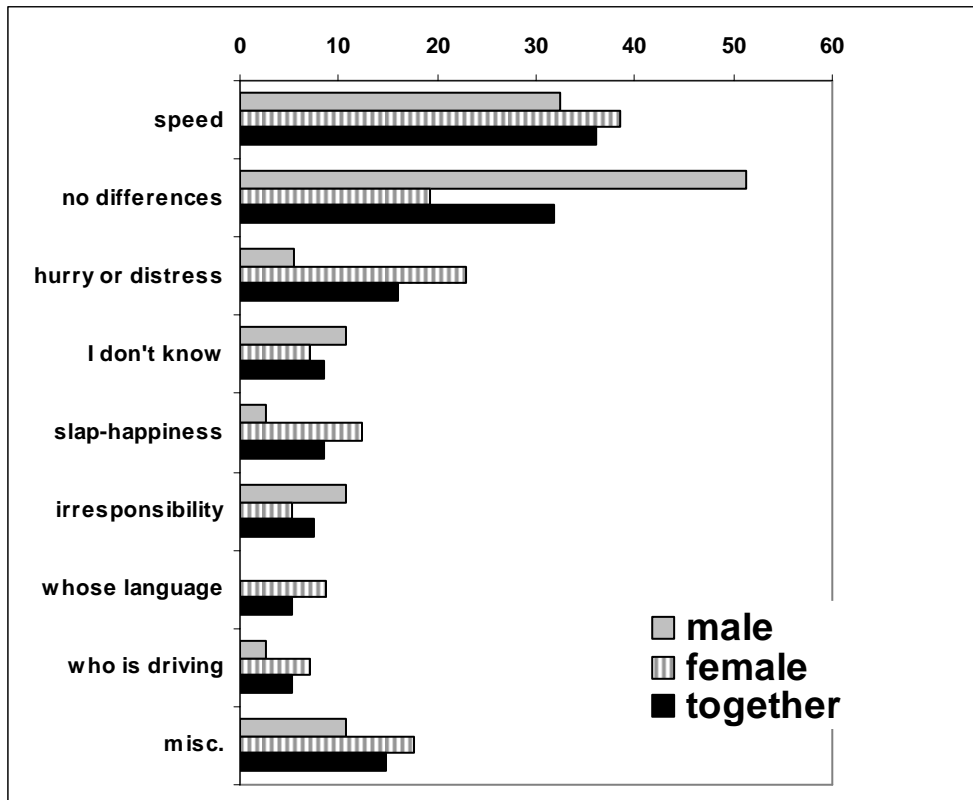


Table 3. Table 3 demonstrates the percentages of informants (separately shown for males, females and both together) who answered that the given feature is important in distinguishing the idioms. Misc. includes cases of playing some role (3), chase (3), driving skills (2), tone of the expression (2), connection other than driving a car (2), danger (1), or empty answer (1).

Most of the answers suggested a difference that is related to the speed (Table 3), while only a small number of informants differentiated the idioms by various other features. An interesting observation in this study

section is that there are marked gender differences. Males gave “no difference” answers more than twice as often as females, whereas females emphasized “hurry or distress” many times more often than males. These observations raise the possibility that the semantics of at least some idioms may not be as gender-independent as was recently suggested for groups of single words (see Vanhatalo 2004).

2.3.4 Methodological considerations

While the overall rationale for using experimental methods in semantic research is well established, there are a few aspects in this particular study setup that deserve more attention.

First, to the best of this author’s awareness, this is the first study of idioms using a methodology of this kind. Therefore the present results needed to serve also as an internal control for the suitability of this approach in idiom research in general. This was accomplished by preselecting the test set of idioms with a pilot study so that there was already some level of confidence concerning the synonymy of many idiom pairs. Hence the actual study was, indeed, bi-directional. While it served to support the hypothesis of synonymy between idioms (the main objective of this paper), it was also testing whether and how well judgements of synonymy may be elicited with this methodology.

Second, it may be pointed out that the idioms studied here might have been too abstract for the present study group of young adults (age range 16–25 yrs.), even though these idioms are all widely known and in everyday use in Finnish society. It is generally thought that the acquisition of highly abstract patterns of language continues long into adulthood. Late acquisition of idiomatic language may be reflected in the *I don’t know* answers, in the few cases with an obvious misunderstanding (e.g. *viilata linssiin* ‘lit. file into lense; pull one’s leg’ > ‘be pedantic,’ obviously caused by the analogous idiom *viilata pilkkua* ‘lit. file a comma; be pedantic’) or unintentional slips (see also Mäntylä and Dufva 2006). It is notable, however, that in the present study design, such a lack of facility in idiomatic language cannot show up as an artifactual synonymy (i.e. show incorrect positive findings), but rather it may lead to an underestimation of the degree of synonymy. The true quantitative figures about synonymy in the wider Finnish-speaking adult population might hence be considerably

higher than the figures obtained in the present paper. Finally, it should be noted that previous studies using a similar test paradigm have shown no significant effect on study results by informant's demographic factors (Vanhatalo 2004), and that even thorough studies have been unable to demonstrate any age-specific maturation in idiom comprehension (e.g. Nippold and Rudzinski 1993; Nippold and Taylor 1995).

Third, in a post-hoc analysis of the open-ended sections it became evident that there are emerging new meanings for several idioms, especially in the younger population. Such multiple meanings do conceivably obscure the clarity (i.e. quantitative level) of synonymy, posing challenges for both testing and glossing.

Fourth, questionnaire tests of this kind are deliberately context-independent. It is possible that, by giving a context for the idioms, the familiarity of the studied idioms would have been much higher, and the descriptions of the idioms in the open-ended tests could have been richer (see also Nippold and Martin 1989). It must be noted, however, that context is also a potential confounder which would be an interesting, easily targeted question for future psycholinguistic studies. Finally, context may not only 'fine tune' but even determine the meanings of idioms, especially in the case of polysemous ones (see also point 3 above).

Fifth, the fixed form type of questionnaire may also lead to errors, such as misconceptions. These would be easily avoided by using an electronic medium.

Sixth, an extensive, multidimensional data matrix, such as the results from this questionnaire test, may be subjected to very different kinds of analyses. This study attempted to find empirical support for the existence of synonymy between idioms; another study could, for instance, search for traces of polysemy in the same results. The full results (data) are always stored in this type of test design, which creates the opportunity to test multiple, even mutually contrasting hypotheses from the same experimental data set.

3. Discussion

By testing native Finnish speakers, the present study provides empirical evidence that synonymy is a distinct and measurable property of idioms. Thinking intuitively, this may feel expected, and it can also be postulated

on theoretical grounds (see Introduction). It is notable, however, that the concept of synonymy between idioms has been little studied in either the linguistic literature or in dictionaries of idioms, leading to a situation where absence of its proof can be easily taken as a proof of its absence. The present study design assumed that the most convincing proof for a lexical variable can be found only from empirical testing of the implicit knowledge of genuine native speakers. As a by-product, the present study design also brought up a multitude of specific features of the tested idioms, that either contribute to their differential meanings or that may imply existing (or emerging) polysemy with some idioms.

In addition to its linguistic findings, this type of study has pragmatic value. The nature of synonymy is, indeed, a challenging topic in current lexicography, especially with regard to the design of (idiom) dictionaries to come. This study was also partly motivated by the two ongoing idiom database projects (Heinonen 2006; Jantunen et al. 2006). While lexicographers are, and will be mainly responsible for collecting and organizing dictionary data, some questions (such as those relating to synonymy) are probably more suitable for studies that are independent of the limitations in time and financial resources that often put pressure on the actual dictionary construction. Such independent (often academic) studies would both support the quality of dictionary content, and could yield commonly available, open-access resources for all relevant future lexicographic projects, whether scientific or commercial. Such co-operative activities are well established worldwide, e.g. in technical and biomedical sciences.

References

- Birih, A. K., Mokienko, V. M. & Stepanova, L. I. (1998/2001) *Slovar' frazeologičeskikh sinonimov russkogo âzyka: okolo 7000 frazeologizmov, svyše 900 sinonimičeskikh râdov*. Moskva: Astrel'.
- Čerkasova T. B. (1991) *Sinonimiâ frazeologičeskikh edinic sovremennogo russkogo âzyka*. Saransk.
- Colston, Herbert (2002) Pragmatic Justifications for Nonliteral Gratitude Acknowledgements. "Oh sure, anytime". *Metaphor and Symbol* 17: 205–226.
- Crystal, David (2003) *A Dictionary of Linguistics and Phonetics*. Malden, MA: Blackwell Pub.
- Danell, Karl J. (1997) Synonymy and Phraseology. In Mariann Skog-Södersved (ed.), *Ethische Konzepte und mentale Kulturen 1. Auf der Suche nach ethischen Konzept-*

- ten in Wortschatz und Phraseologie*, pp. 11–16. Acta Universitatis Umensis, Umea Studies in the Humanities 139. Umeå: Umea University Press.
- Edmonds, Philip (1999) *Semantic Representation of Near-Synonyms for Automatic Lexical Choice*. PhD Thesis, published as technical report CSRI-399, Department of Computer Science, University of Toronto.
- Edmonds, Philip & Hirst, Graeme (2002) Near-synonymy and lexical choice. *Computational Linguistics* 1:1: 1–40.
- Heinonen, Tarja Riitta (2006) Variation and Flexibility within Idiomatic Expression in Finnish. In Marja Nenonen (ed.), *Collocations and Idioms 1: The First Nordic Conference on Syntactic Freezes. Joensuu, Finland, May 19-20, 2006. Abstracts*. Joensuu: University of Joensuu.
- Hümmer, Christiane (2004) A Corpus Based Approach to Near Synonymy of German Multi-Word Expressions. In Petr Sojka & Karel Pala & Pavel Smrz (eds.), *Proceedings of the Second International WordNet Conference*. Brno: Masaryk University.
- Hyvärinen, Irma (1996) Zur Semantik von deutschen und finnischen Verbidiomen. In Jarmo Korhonen (ed.) *Studien zur Phraseologie des Deutschen und des Finnischen II*, pp. 345–439. Studien zur Phraseologie und Parömiologie 10. Bochum: Universitätsverlag Brockmeyer.
- Jantunen, Jarmo H. & Kokko, Ossi & Muikku-Werner, Pirkko (2006) Towards building a Finnish corpus-based idiom dictionary. In Marja Nenonen (ed.), *Collocations and Idioms 1: The First Nordic Conference on Syntactic Freezes. Joensuu, Finland, May 19–20, 2006. Abstracts*. Joensuu: University of Joensuu.
- Kari, Erkki (1993) *Naulan kantaan. Nykysuomen idiomisanakirja*. Helsinki: Otava.
- Korhonen, Jarmo (1995) *Studien zur Phraseologie des Deutschen und des Finnischen I*. Studien zur Phraseologie und Parömiologie 7. Bochum: Universitätsverlag Brockmeyer.
- (2001) *Alles im Griff. Homma hanskassa. Saksa-suomi-idiomisanakirja. Idiomwörterbuch Deutsch-Finnisch*. Helsinki: WSOY.
- Kunin, A. V. (1996) *Kurs frazeologii sovremennogo anglijskogo äzyka*.
- Mäntylä, Katja & Dufva, Hannele (2006) Slippery idioms. In Marja Nenonen (ed.), *Collocations and Idioms 1: The First Nordic Conference on Syntactic Freezes. Joensuu, Finland, May 19–20, 2006. Abstracts*. Joensuu: University of Joensuu.
- Nenonen, Marja (2002) *Idiomit ja leksikko. Lausekeidiomien syntaktisia, semanttisia ja morfologisia piirteitä suomen kielessä*. Joensuun yliopiston humanistisia julkaisuja 29. Joensuu: University of Joensuu.
- Nippold, Marilyn A. & Martin, Stephanie Tarrant (1989) Idiom Interpretation in Isolation versus Context. A Developmental Study with Adolescents. *Journal of Speech and Hearing Research* 32: 59–66.
- Nippold, Marilyn A. & Rudzinski, Mishelle (1993) Familiarity and Transparency in Idiom Explanation. A Developmental Study of Children and Adolescents. *Journal of Speech and Hearing Research* 36: 728–737.

- Nippold, Marilyn A. & Taylor, Catherine L. (1995) Idiom Understanding in Youth. Further Examination of Familiarity and Transparency. *Journal of Speech and Hearing Research* 38: 426–433.
- Nirvi, R. E. (1944) *Sanankieltoja ja niihin liittyviä kielenilmiöitä itämerensuomalaisissa kielissä. Riista- ja kotieläintalous*. Suomalaisen Kirjallisuuden Seuran Toimituksia 223. Helsinki: SKS.
- Penttilä, Esa (2006) *It Takes an Age to Do a Chomsky: Idiomaticity and Verb Phrase Construction in English*. Joensuu: University of Joensuu.
- Rapola, Martti (1944) *Merkitysoipillinen väitöskirja itämerensuomalaisten kielten alalta. R. E. Nirvi: Sanankieltoja ja niihin liittyviä kielenilmiöitä itämerensuomalaisissa kielissä. Riista- ja kotieläintalous. SKST 223. Virittäjä* 48: 182–187.
- Raukko, Jarno (1999) An “intersubjective” method for cognitive-semantic research on polysemy: The case of GET. In Masako K. Hiraga & Chris Sinha & Sherman Wilcox (eds.), *Cultural, psychological and typological issues in cognitive linguistics. Selected papers of the bi-annual ICLA meeting in Albuquerque, July 1995*, pp. 87–105. Current Issues in Linguistic Theory 152. Amsterdam/Philadelphia: John Benjamins.
- Rekiaro, Päivi (1998) *101 idiomia in English*. Jyväskylä: Gummerus.
- Soshnikova, Nata (2006) Synonymity of phraseological units with the component ‘give’. In Marja Nenonen (ed.), *Collocations and Idioms 1: The First Nordic Conference on Syntactic Freezes. Joensuu, Finland, May 19–20, 2006. Abstracts*. Joensuu: University of Joensuu.
- Tuomola, Urho (1935) Sanojen merkitysten huomioimisesta kansankielen sanaston keuruissa. *Virittäjä* 39: 137–145.
- Vanhatalo, Ulla (2004) Koehenkilöiden taustan merkitys kokeellisessa leksikaalisessa semantiikassa ja sen sovelluksissa. *Sananjalka* 46: 81–105.
- (2005) *Kyselytestit synonymian selvittämisessä. Sanastotietoutta kielenpuhujilta sähköiseen sanakirjaan*. Helsinki: Yliopistopaino.
- (2006) Vedäkö nenästä vai viilaanko linssiin? Voiko idiomeilla olla synonymiaa? In Piia Seppälä (ed.), *XXXIII Kielitieteen päivät Turussa 4.–5. toukokuuta 2006. Esitelmien tiivistelmät*, p. 122. Turku: University of Turku.
- Varis, Markku (1998) *Sumea kieli. Kiertoilmausten muoto ja intentio nykysuomessa*. Suomalaisen Kirjallisuuden Seuran Toimituksia 717. Helsinki: SKS.
- Virkkunen, Sakari (1974) *Suomalainen fraasisanakirja Kivestä Kekkoseen*. Helsinki: Otava.
- Žukov, V. P. (1987) *Slovar’ frazeologičeskikh sinonimov russkogo ōzyka*. Moskva: Russkij ōzyk.

Contact information:

Ulla Vanhatalo
Department of General Linguistics
P.O. Box 9
FIN-00014 University of Helsinki
email: ulla(dot)vanhatalo(at-sign)helsinki(dot)fi

Appendix 1

List of the studied Finnish idioms with their literal meaning and approximate English translation. The numbers show how the idioms were grouped. The idioms were formed to be understandable by adolescents (e.g. *ei ymmärtää hölkäsen pöläystä* instead of *jku ei ymmärrä jstak hölkäsen pöläystä*).

in Finnish	literally	meaning
1. ajaa kaasu pohjassa	drive gas in the bottom	drive very fast
1. ajaa nasta laudassa	drive pin in the blank	drive very fast
1. ajaa tuhatta ja sataa	drive thousand and hundred	drive very fast
2. antaa kenkää	give shoe	lay off
3. ei ymmärtää hölkäsen pöläystä	not understand *	not understand
3. ei ymmärtää tuon taivaallista	not understand heavenly	not understand
4. haistella ilmaa	smell air	sound out
5. heittää hanskat tiskiini	throw gloves on the desk	give up
2. heittää pellolle	throw to the field	lay off
6. hyppiä nenälle	jump onto nose	be forward
7. hävitä kuin tuhka tuuleen	disappear as ash into wind	disappear without
8. juosta kieli vyön alla	run lip under belt	be in a great hurry
8. juosta pää kolmantena jalkana	run as head the third leg	be in a great hurry
7. kadota kuin maan alle	disappear as under the ground	disappear without trace
9. kaivaa verta nenästään	delve blood from one's nose	irritate
9. kerjätä turpaansa	beg for muzzle	irritate
10. polttaa hihansa	burn one's sleeves	flare up
10. repiä pelihousunsa	rip one's game pants	flare up
6. tallata varpaille	stamp on toes	be forward
11. temmata hatusta	pull from hat	guess
4. tunnustella maaperää	explore ground	sound out
11. vetää hihasta	pull from sleeve	guess
12. viilata linssiin	file into a lens	cheat
5. viskata kirves kaivoon	throw axe into well	give up
12. vetää nenästä	pull nose	cheat
'Fillers:'		
13. ahtaa kitaansa	push one's mouth	wolf down
13. heittää yrjöt	throw georges	throw up
14. nyrpistää nenäänsä	purse one's nose	be supercilious
15. olla ihan puhki	be totally through	be really tired
16. olla mieli mustana	be with black mind	be sad
16. olla tuli hännän alla	be with fire under tail	be nervous or in
14. ottaa nokkiinsa	take into one's noses	be provoked
15. saada hepuli	to get a seizure	get upset