by subordinate clauses where the subordinating conjunction or the relative pronoun may be regarded as the head.

We will not discuss these problems further but hope that the terms used in *Segram* should be acceptable or at least stimulate discussions of grammatical terminology. Students have long requested better grammatical terminology reflecting a unified system which is easier to understand.

We are grateful to Arthur Holmer, Lund for comments on this paper.

References

Chomsky, Noam. 1957. *Syntactic structures.* The Hague: Mouton.
Diderichsen, Paul. 1946. *Elementær dansk grammatik.* København: Gyldendal.
Lastow, Birgitta & Gisela Håkansson. 1997. 'Grammatical terminology and the application *Gramte*'. *Working papers* 46, 185-95. Dept of Linguistics, Lund University.
Pienemann, Manfred. 1998. *Language processing and second language development: processability theory.* Amsterdam: John Benjamins.
Pienemann, Manfred & Gisela Håkansson. 1999. 'A unified approach toward the development of Swedish as L2: a processability account'. *Studies in second language acquisition* 21, 383-410.
Sigurd, Bengt (ed.). 1994. *Computerized grammars for analysis and machine translation.* Lund: Lund University Press.
Svenska Akademiens grammatik (SAG). 1999. Stockholm: Norstedts.

Bengt Sigurd <Bengt.Sigurd@ling.lu.se>
Gisela Håkansson <Gisela.Hakansson@ling.lu.se>
Johan Dahl <Johan.Dahl@ling.lu.se>

Lund University, Dept. of Linguistics
Working Papers 51 (2005), 187–198

Machine translation of marine forecasts, quarterly company reports and recipes between Swedish, English, Malay and Chinese

Bengt Sigurd, Birgitta Lastow, Hong Gao and Mats Eeg-Olofsson

Abstract

This paper will report on four machine translation projects dealing with restricted domains and controlled or semi-controlled language. The computer program *Seatra* translates marine forecasts between Swedish and English. It is in daily use by the Swedish meteorological agency (SMHI) and generally needs no postediting. The same basic program but with different lexicons is used in the program *Maltra* which can translate official Malaysian marine forecasts into English and Chinese.
The program *Reptra* dealing with quarterly company reports and *Receptra* dealing with recipes use the same platform. All these programs (which can be used bidirectionally) will be presented and commented on, above all from a linguistic point of view. It is suggested that no more than 3 mistakes per 100 words be accepted if the postediting should not be too heavy, and this quality can be obtained in restricted-domain systems.

1 *Seatra*

The Canadian program *METEO* (Kittredge et al. 1973; Goldberg, Kittredge & Polguère 1988) is well known and the task to translate marine forecasts from Swedish into English suggested by the Swedish meteorological agency SMHI didn't seem too hard. Early work is described in Sigurd, Lastow & Vävargård 1996. The program *Seatra* was developed in Prolog (LPAProlog http://www.lpa.co.uk), using a technique which is based on lexicons containing single words, multi-word phrases, discontinuous phrases and some grammatical phrases such as noun phrases, prepositional phrases and adverbial phrases. The program is designed to produce simple and fast translation in a restricted domain without offering deep syntactic and

semantic analyses. The *Seatra* lexicon includes about 2000 items. *Seatra* is an interlingua system, where word meanings are rendered in a special format, but certain simple transformations are used in the generation process in order to get the right word order and agreement in English.

The SMHI marine forecasts (see below) typically include a part where the characters and movements of the synoptic objects (lows, highs and fronts) are mentioned (cf. Sigurd et al. 1992). It may also include warnings. Then the weather in the different regions (*Skagerrak, Kattegat, Vänern*, etc) is mentioned. The synoptic part causes the main part of the difficulties to an automatic translation system as it includes finite sentences which may require changed word order and agreement in English. Some agreement problems can be overcome by using telegraphic style instead of full style (*A high is approaching Scandinavia* > *High approaching Scandinavia*), which is an option in the program.

The following is an example of the translation of a Swedish marine forecast by *Seatra*. The wind speed given in m/s in Swedish is not converted into knots often used in English forecasts.

Utfärdad 000218 kl 13:00. Utsikter till Lördag kväll.

VÄDERÖVERSIKT. En högtrycksrygg täcker Skandinavien. Över västra Nordsjön finns ett lågtryck som är på väg åt sydost och befinner sig i morgon middag över Tyskland.

KULINGVARNING ÄR UTFÄRDAD FÖR Yttre Skagerack.

DETALJPROGNOS.
SKAGERACK, KATTEGATT. Syd ca 7 m/s ökande, från i kväll sydost 8-12, på yttre Skagerack lokalt kuling 15. I morgon ost eller nordost sakta avtagande. Mitt på dagen 6-9. Måttlig till god sikt, tidvis nederbörd.

VÄNERN. Sydost 2-6, i morgon nordost. Mest god sikt.

ÖRESUND OCH BÄLTEN, SYDVÄSTRA och SÖDRA ÖSTERSJÖN. Sydost 6-9, ökande till 8-13, som mest i väster. I morgon eftermiddag ost och långsamt avtagande. Måttlig sikt med nederbörd tidvis.

SYDÖSTRA, MELLERSTA OCH NORRA ÖSTERSJÖN. Sydväst 2-6, från i kväll sydost och söder om Gotland något ökande. Måttlig till god sikt.

ÅLANDSHAV OCH SKÄRGÅRDSHAVET, BOTTENHAVET, NORRA KVARKEN, BOTTENVIKEN. Mest västlig vind 2-7 m/s, från i natt sydväst. God sikt men på Bottenviken till en början måttlig sikt.

KULINGVARNING FÖR YTTRE SKAGERACK. I natt ost eller sydost lokalt 15 m/s.

Machine translation by *Seatra*:

Issued 218 time 13:00. Forecast until Saturday evening.

WEATHER SURVEY.
A ridge of high covers Scandinavia. Over western North sea a low occurs which is on the way southeastwards and is staying tomorrow noon over Germany.

GALE WARNING IS ISSUED FOR outer Skagerrak.

DETAILED FORECAST.
SKAGERRAK, KATTEGAT. Southerly ca 7 m/s increasing, from the evening southeasterly 8-12, at outer Skagerrak locally near gale 15. Tomorrow easterly or northeasterly slowly decreasing. In the middle of the day 6-9. Moderate to good visibility, at times precipitation.

LAKE VAENERN. Southeasterly 2-6, tomorrow northeasterly. Mainly good visibility.

THE SOUND AND THE BELTS, SOUTHWESTERN AND SOUTHERN BALTIC. Southeasterly 6-9, increasing to 8-13, most in western waters. Tomorrow afternoon easterly and slowly decreasing. Moderate visibility with precipitation at times.

SOUTHEASTERN, CENTRAL and NORTHERN BALTIC. Southwesterly 2-6, from the evening southeasterly and south of Gotland somewhat increasing. Moderate to good visibility.

SEA OF AALAND AND AALAND ARCHIPELAGO, SEA OF BOTHNIA, THE QUARK, BAY OF BOTHNIA. Mainly westerly wind 2-7 m/s, from tonight southwesterly. Good visibility but at Bay of Bothnia at first moderate visibility.

GALE WARNING FOR OUTER SKAGERRAK. Tonight easterly or southeasterly locally 15 m/s.

1.1 Comments

This English translation (some 200 words) shows no serious mistakes. The choice of vocabulary is generally straightforward but sometimes causes problems. Ambiguities are not as problematic as some persons think when they express their pessimism about the potential of machine translation. Many ambiguities are resolved by the syntactic rules of the parsing system. Ambiguities can also be resolved by the domain characteristics or frequency

information. The word *var* can be the equivalent of the question word *where*, but questions do not occur in the texts of the domains treated in our four projects. The word *var* can also denote 'pus', but this a very rare concept, which only appears in special texts. In the weather texts, *var* is a verb and the equivalent of *was/were*.

The Swedish word *någon* can mean 'somebody' but this meaning does not occur in the texts studied. The word causes problems, however, in phrases such as: *någon regnskur*, which is a singular noun phrase. It can be translated by *some rain showers*, but *some* is plural and the noun then has to be changed into plural. In *Seatra, någon* is translated in this context by *an odd (rain shower)* which is a way of keeping the number of the noun.

A human translator may sometimes have chosen other ways of expression, e.g. *on its way* instead of *on the way*. The system is also programmed to express *vridande* 'turning' alternatively by the special terms *veering* 'clockwise' and *backing* 'counter-clockwise'. The translation of *vridande* in the phrase *sydost mot kvällen ökande och vridande till syd* requires taking the previous wind direction *(sydost)* into account to get *veering*. Words not found in the lexicon and misspelled words go right through the system which thus serves as a spell checker for Swedish.

As can be seen, the translation is very good, although the word order, especially the placement of adverbs is sometimes a bit strange. A general problem is the Swedish inverted word order after an initial adverbial phrase as illustrated by: *Under morgondagen rör sig ett lågtryck mot Finland/ Tomorrow (a) low (is) moving towards Finland.* The prepositional phrase *mot Finland* may be a post-modifier of *lågtryck* (low) or an adverbial of the verb *rör sig*. Due to the inversion used in Swedish, the phrase *mot Finland* occurs after *lågtryck* but *ett lågtryck mot Finland* is a strange noun phrase and *mot Finland* has to be taken as an adverbial of *rör sig*.

2 *Maltra*

It is interesting to test the translation platform on different languages and thanks to the assistance of Gao Hong (Chinese), Niclas Burenhult (Malay) and Arthur Holmer (Chinese, Malay) a system translating between Malay, English and Chinese was developed. The lexicons were changed accordingly. Official forecasts are published by the Malaysian government and found at http://www.kjc.gov.my/english/weather/forecast/ship.html. The web site also publishes an English version, thus allowing comparison between it and our Maltra version. It is not known how the Malaysian version is produced, but

our *Maltra* machine translation is as good and gives the right order of the adjective and the noun in noun phrases (see *cuaca baik* 'fair weather' below).

As can be seen the official Malaysia forecasts given here for four days are much shorter than the Swedish forecasts. They do not include a synoptic part treating lows and highs, and the regional weather gives somewhat different information. The Malaysian government's forecasts give information for the morning, afternoon and night about clouds, precipitation, wind direction, wind speed and sometimes wave heights and warnings.

The following is the Malay original (somewhat edited) and English and Chinese translations by *Maltra*.

Ahad 21/07/2002
Pagi: Cuaca baik.
Petang: Hujan curah di satu dua tempat.
Malam: Mendung.
Angin: BD 20-30 km/j.
Ombak: 1.0-1.5 meter.
Hujan curah di satu dua tempat
Petang.

Isnin 22/07/2002.
Pagi: Mendung.
Petang: Cuaca baik.
Malam: Cuaca baik.
Angin: BD 25-30 km/j.
Ombak: 1.0-1.5 meter.
Mendung
Pagi.

Selasa 23/07/2002.
Pagi: Cuaca baik.
Petang: Cuaca baik.
Malam: Cuaca baik.
Angin: BD 20-30 km/j.
Ombak: 1.0-1.5 meter.
Cuaca baik
Sepanjang Hari.

Rabu 24/07/2002.
Pagi: Ribut petir di satu dua tempat.
Petang: Cuaca baik.
Malam: Cuaca baik.
Angin: BD 20-30 km/j.
Ombak: 1.0-1.5 meter.
Ribut petir di satu dua tempat
Pagi.

English translation by *Maltra*

Sunday 21 / 7 / 2002
morning : fair weather .
afternoon : scattered showers .
night : overcast .
wind : SW 20 - 30 km/h .
wave height : 1 - 1.5 meter .
scattered showers afternoon .

Monday 22 / 7 / 2002 .
morning : overcast .
afternoon : fair weather .
night : fair weather .
wind : SW 25 - 30 km/h .
wave height : 1 - 1.5 meter .
overcast morning .

Tuesday 23 / 7 / 2002 .
morning : fair weather .
afternoon : fair weather .
night : fair weather .
wind : SW 20 - 30 km/h .
wave height : 1 - 1.5 meter .
fair weather all day .

Wednesday 24 / 7 / 2002 .
morning : thunderstorms locally .
afternoon : fair weather .
night : fair weather .
wind : SW 20 - 30 km/h .
wave height : 1 - 1.5 meter .
thunderstorms locally morning .

Chinese translation of the same forecast by *Maltra*:

星 期 日 21 / 7 / 2002
上 午 : 晴 朗 的 天 气 .
下 午 : 分 散 的 阵 雨 .
夜 晚 : 多 云 .
风 : 西 南 20 - 30 公 里 每 小 时 .
波 浪 高 度 : 1 - 1.5 公 尺 .
分 散 的 阵 雨 下 午 .

星 期 一 22 / 7 / 2002 .
上 午 : 多 云 .
下 午 : 晴 朗 的 天 气 .
夜 晚 : 晴 朗 的 天 气 .
风 : 西 南 25 - 30 公 里 每 小 时 .

波 浪 高 度 : 1 - 1.5 公 尺 .
多 云 上 午 .

星 期 二 23 / 7 / 2002 .
上 午 : 晴 朗 的 天 气 .
下 午 : 晴 朗 的 天 气 .
夜 晚 : 晴 朗 的 天 气 .
风 : 西 南 20 - 30 公 里 每 小 时 .
波 浪 高 度 : 1 - 1.5 公 尺 .
晴 朗 的 天 气 全 天 .

星 期 三 24 / 7 / 2002 .
上 午 : 雷 暴 雨 当 地 .
下 午 : 晴 朗 的 天 气 .
夜 晚 : 晴 朗 的 天 气 .
风 : 西 南 20 - 30 公 里 每 小 时 .
波 浪 高 度 : 1 - 1.5 公 尺 .
雷 暴 雨 当 地 上 午 .

2.1 Comments

The translation of the Malaysian marine forecasts into English and Chinese is fairly straightforward and the problems are mainly of a lexical nature. In Malay, however, the noun phrase rule is different, as the attributive adjective occurs after the head noun as is illustrated in *cuaca baik* 'fair weather', where *cuaca* means 'weather'. Both Malay and Chinese lack agreement.

3 *Receptra*

Recipes are a special genre and anyone interested in cooking knows how recipes are generally written and what words and phrases are used (cf. Anward 1994). Recipes belong to the genre called manuals/instructions which has interested researchers in machine translation for a long time, as translation of manuals is of great commercial interest. The program *Receptra* is based on the same platform as *Seatra* but with an extended vocabulary and set of phrases. The vocabulary includes names of courses, products, materials, etc. Furthermore the program includes a great number of words for processes (e.g. frying, boiling, peeling, baking, pouring, melting) which is a characteristic of this domain.

The data for *Receptra* are some 50 recipes taken from cookbooks, journals and newspapers. The lexicon of *Receptra* includes about 4,000 items. The following shows typical recipes and automatic translations by *Receptra*.

Swedish recipe (after the ingredients have been listed):

> Sätt ugnen på 200 grader. Riv potatisen fint. Blanda alla ingredienser till degen. Klä en pajform med degen. Förgrädda pajskalet cirka 10 minuter. Tina spenaten och skär den grovt. Om du använder färsk spenat, koka den någon minut och låt rinna av. Fördela spenaten i pajskalet. Lägg tunna laxskivor som ett täcke över. Vispa ihop ägg, mjölk, ost, ingefära, salt och peppar. Häll det över pajen. Grädda mitt i ugnen i cirka 30 minuter.

English translation by Receptra:

> Set the oven at 200 degrees. Grate the potatoes finely. Mix all ingredients to the dough. Dress a pie form with the dough. Pre-bake the pie shell circa 10 minutes. Melt the spinach and cut it rough. If you use fresh spinach, boil it an odd minute and let drain. Spread the spinach in the pie shell. Put thin salmon slices as a cover over. Beat together egg, milk, cheese, ginger, salt and pepper. Pour it over the pie. Bake in the middle of the oven in circa 30 minutes.

The English translation is reasonably good, although some sentences may sound strange (e.g. *Put thin salmon slices as a cover over*). We note all the ingredients necessary in machine translation systems; some may be difficult to identify in other cultures.

Swedish recipe (after the ingredients have been mentioned):

1. Skala potatis och selleri.
2. Strimla purjolöken och fräs den mjuk i smöret. Häll av potatisvattnet och mosa med en stöt direkt i kastrullen.
3. Rör ner purjolöken och lite mjölk i taget. Krydda med salt och peppar. Rör moset luftigt.
4. Dela svampen och fräs den i smör några minuter. Krydda med salt, peppar och persilja.
5. Dela baconskivorna i halvor och stek dem knapriga. Låt dem rinna av på hushållspapper.
6. Krydda köttet med salt och peppar och stek det i smör ca 15 minuter på varje sida.
7. Servera köttet på moset med svamp och bacon till.

Automatic translation by *Receptra*:

1. Peel potatoes and celery.
2. Shred the leek and fry it soft in the butter. Pour the potato water away and mash with a pestle directly in the saucepan.
3. Stir the leek down and some milk a little at a time. Season with salt and pepper. Stir the mash light.

4. Divide the mushroom and fry it in butter some minutes. Season with salt, pepper and parsley.
5. Divide the bacon slices in halves and fry them crisp. Let them drain at kitchen rolls.
6. Season the meat with salt and pepper and fry it in butter ca 15 minutes at each side.
7. Serve the meat at the mash with mushroom and bacon.

3.1 Comments

We note the imperative verb forms: *set, grate (finely), mix, dress, pre-bake, melt, cut rough, let drain, boil, spread, put, pour, bake*. Some recipes include particle verbs which may cause translation problems: *häll av* 'pour away', *smaka av* 'season'. A dictionary suggests translating *koka upp* by 'bring to the boil' and *koka sönder* by 'boil to shreds'. The system needs special procedures in order to identify particle verbs, which are treated as discontinuous lexical entries. Problems may appear when there is another constituent (e.g. an adverbial) between the verb and the particle. This is illustrated by *häll försiktigt över såsen i en kastrull*, which should be translated 'pour the sauce carefully in a saucepan' and not 'pour carefully over the sauce in a saucepan'. The problem can be solved by using discontinuous lexical items.

Swedish pronouns (*den, det*) constitute problems. Similarly adjective forms in *-t* cause problems as they can be neuter forms of adjectives or adverbial forms. The measure units (*g, hg, knivsudd* 'knife's point', *nypa* 'pinch') may cause problems. Modern Swedish does not use *uns* for 'oz'. A good recipe translation system should be able to convert units of measurement.

We note the phrase *(lite) i taget*, which may be translated by the fixed phrase '(a little) at a time'. The sentence *Servera köttet på moset med svamp och bacon till* seems to contain a discontinuous phrase *med ... till*, which can best be translated by 'with' only.

4 *Reptra*

Automatic English translation of financial texts is of great commercial interest in the globalized world. Quarterly company reports seem to constitute a fairly restricted domain and it is interesting to investigate how well automatic translation of Swedish quarterly reports works.

The following is an example of the quarterly reports published in Swedish financial or news media:

Vinst för Advise.
Advise, som noteras på Nya Marknaden, redovisar ett resultat efter finansnetto på 0,5 miljoner kronor för tremånadersperioden maj-juli 1999 (0,9). Rörelseresultatet blev 0,5 miljoner kronor (-1,1). Nettoomsättningen uppgick till 7,2 miljoner kronor (4,1).

English translation by *Reptra*:

Profit for Advise.
Advise, which is registered at Nya Marknaden, reports a result after financial items of SEK 0.5 million for the three months period May-July 1999 (0.9). The operative profit was SEK 0.5 million (-1.1). The net turnover amounted to SEK 7.2 million (4.1).

Note that Swedish decimal commas have to be changed into decimal points. The following is a second example of quarterly reports:

Försämrad vinst för Måldata.
IT-bolaget Måldata redovisar en vinst före skatt på 7,8 miljoner för helåret 1999. Det är en försämring jämfört med året innan då vinsten uppgick till 20,1 miljoner. Vinstförsämringen beror enligt bolaget på en försvagning i marknaden för affärssystem på grund av millennieskiftet. Omsättningen ökade från 277 miljoner till 306 miljoner. Styrelsen föreslår en utdelning på 60 öre per aktie, jämfört med 2:25 föregående räkenskapsår.

English translation by *Reptra*:

Decreased profit for Måldata.
The IT company Måldata reports a profit before tax at 7. 8 million for the whole year 1999. It is a decrease compared to the year before då the profit amounted to 20. 1 million. The decrease of profit is caused according to the company at a weakening in the market for business system because of the millennium shift. The turnover increased from 277 million to 306 million. The board suggests a dividend at 60 öre per share, compared to 2:25 preceding budget year.

The program *Reptra* is based on some hundred quarterly reports of different lengths. The content is rather similar (profit, turnover, problems (cuts), expansion, development, shares, dividends, prospects). The verbs are typically: *changed*, *increased*, *decreased*, *fell*, *rose*, *caused*, *amounts to*, *depends on*.

The translation above shows that the Swedish word *då* 'when' is not in the lexicon. The verb *beror på* in the Swedish text *Vinstförsämringen beror enligt bolaget på* has not been translated correctly as the computer has not identified the verb + preposition *beror ... på* 'depends on' (cf. the problem of particle verbs treated above). The Swedish word *affärssystem* may be singular or plural. The computer has not managed to make the correct decision.

The translation of currency (crowns, pounds, dollars, etc) is a problem as the order between the amount and the currency has to be changed in some cases (amount + kronor > SEK + amount.).

5 Conclusions

The projects show that it is possible to build an automatic translation system with 100 percent translation quality if the domain and the language is completely restricted. But for each degree of added information and stylistic freedom the quality risks becoming less perfect. As a rough measure one may state that 1 mistake in 100 words is acceptable as it does not disturb a reader much and can easily be corrected by a posteditor. But more than 3 mistakes is unacceptable as it irritates the reader and makes a posteditor want to translate the text himself without the help of the computer.

Maltra can produce good translations, but the texts are the least varied and shortest. *Seatra* produces perfect translations although the texts can be quite long and informative. *Seatra* is known to have the positive effect that meteorologists are eager to keep to the standard language in order to avoid having to postedit the text.

The automatic translations by *Receptra* show that it is possible to produce good translations of recipes, if the ingredients and procedures are not too exotic. But writers of recipes often want to show individual styles and are probably not generally willing to obey strict rules. *Receptra* can help in spreading recipes and cooking customs between the different cultures of the world.

Reptra can achieve good translation if the style is not too varied and no additional topics such as international events, new products, new taxes, etc are mentioned. It is probable that the international business world accepts nothing less than perfect translations of quarterly reports. This can hardly be guaranteed by machine translation systems and therefore careful human postediting will be necessary.

Only a few domains (weather, recipes, special manuals, quarterly reports, stock market reports, life stock market reports, travel agency dialogue, job advertisements, route descriptions) have been investigated with a view to automatic translation. It is a challenge to find further restricted domains which can be translated by automatic systems of commercial interest.

References

Anward, Jan. 1994. 'Från *potatisen kokas* till *koka potatisen*' In *Språkbruk, grammatik och språkförändring*. Lund: Institutionen för nordiska språk.

Goldberg, Eli, Richard Kittredge & Alain Polguère. 1988. 'Computer generation of marine weather forecasts'. *Journal of atmospheric and oceanic technology* 5:4, 472-83.

Kittredge, Richard et al. 1973. *TAUM-73*. Montréal: Université de Montréal.

Sigurd, Bengt, Mats Eeg-Olofsson, Caroline Willners & Christer Johansson. 1992. 'Deep comprehension, automatic translation and generation of weather reports (*Weathra*)'. *Proceedings of the Coling* 92, Nantes, Vol II, 749-55.

Sigurd, Bengt, Birgitta Lastow & Tomas Vävargård. 1996. 'Computer generation of weather overviews'. Paper presented at The European conference on artificial intelligence (ECAI-96), Budapest.

Bengt Sigurd <Bengt.Sigurd@ling.lu.se>
Birgitta Lastow <Birgitta.Lastow@ling.lu.se>
Hong Gao 高虹, University of Toronto <hgao@utm.utoronto.ca>
Mats Eeg-Olofsson <Mats.Eeg-Olofsson@ling.lu.se>

Lund University, Dept. of Linguistics
Working Papers 51 (2005), 199–207

Basic verb frequency in Megrelian

Revaz Tchantouria and Karina Vamling

Our aim is to investigate which verbs show the highest textual frequencies in the Kartvelian language Megrelian. The general assumption is that unmarked verbs represent lexical core concepts and that they will emerge among the verbs with the highest text frequencies, showing crosslinguistic similarities.

1 Background

1.1 Megrelian

Megrelian is a Kartvelian (South Caucasian) language spoken in Western Georgia. The number of speakers is estimated to approximately half a million (no official census data are available, as Megrelians are registered as Georgians). Apart from a short period around the 1930s, when some attempts were made to standardize the language (Vamling 2000), the language is not used in writing. Speakers are generally bilingual, using Georgian as their literary language for all administrative and educational purposes.

1.2 Basic verbs and text frequency[1]

Unmarked verbs are general and basic and are assumed to represent lexical core concepts and appear among the verbs with the highest text frequencies. When compared in various languages they also show patterns of cross-linguistic regularities. Typically, such verbs – *see, say, take, go, know* – form the nucleus of semantic fields: Perception, Verbal Communication, Possession, Motion, Cognition (Viberg 1994).

Such a concept of markedness is central to typological research on hierarchies and prototypes (following Greenberg 1966) and allows not only for binary relations but extends to several values along a hierarchy of relative markedness. An example is the hierarchy of sense modalities, which are ranked from unmarked to higher degree of markedness: *sight* > *hearing* >

[1]This study on verb frequency in Megrelian has been undertaken in connection with research on cross-linguistic lexicology headed by professor Åke Viberg at the Department of Linguistics, Lund Univerisity. An earlier version of the paper was presented at the Chikobava Conference in Tbilisi in 1998.