

NIVERSITY OF GOTHENBURG ept of swedish The Institute for Swedish as a Second Language The Centre for Lexicology and Lexicography The Swedish Language Bank

A Swedish Academic Word List: Methods and Data

Håkan Jansson hakan.jansson@gu.se Sofie Johansson Kokkinakis sofie.johansson.kokkinakis@gu.se

Judy Ribeck judy.ribeck@gu.se Emma Sköldberg emma.skoldberg@gu.se

Aim of the project

The purpose is to construct a Swedish academic word list that can be used in language teaching.

It will also be interesting from a lexicological perspective as a mapping of the Swedish academic vocabulary.

Inspirational background

Since its release, *The Academic Word List* (Coxhead 2000, AWL) has hugely influenced the curricula of English for academic purposes and English as a second/foreign language.

Summary

Academic language often presents a challenge to students, both language learners and native speakers. Therefore there is a need for educational language tools such as academic vocabulary resources.

This poster reports on different approaches to compiling a corpus of authentic academic text material. The purpose is to compose an empirical basis for the construction of a Swedish academic word list. In analogy with the compilation of *The Academic Word List* (Coxhead 2000), the corpus content is crucial to the final content of our word list.

Extraction of academic words

A great deal of higher education in Sweden is in English and almost 90% of all Ph.D. theses are written in English. The Language Council of Sweden expresses a wish for Swedish researches to be able to communicate their results in their mother tongue.

Possible methods and data

- 1. Translation of AWL (manual or automatic)
- 2. Collection of documents from universities by WebBootCat
- 3. Manual dowloading of publications from a national academic online database (SwePub)

	+	
1.	(well) tried method (cf. Awl of Portuguese)	Problems related to polysemy and homonymy; Representative Swedish academic words?
2.	A large and quickly created corpus	Data needs to be structured; Only some university subjects are represented; Academic language?
3.	Possibilty to balance the corpus w.r.t. subjects, text types etc.	Size limited due to the restricted selection of texts

Automatic identification of words with the following characteristics:

- High frequency and even dispersion in the texts of the six university subjects
- Higher frequency in the academic texts than in other text types, e.g. novels
- Not part of the 1000 most frequent words of everyday language of Swedish

Result: 750 words

	The top 20 art words									
	dock relation	6. begrepp 7. form	11. perspektiv 12. beskriva	16. utgöra 17. grund						
	samt	8. betydelse	13. text	18. såväl						
4.	studie	9. analys	14. innebära	19. enligt						
5.	Social	10. utifrån	15. exempelvis	20. därmed						

Future plans

• Add information about the words (part-of-speech, inflection, meaning, language examples, English translations)

The Swedish academic corpus of arts

	Ph.D. theses	Articles	Other	Total	
1. Ethnology	1,210,735	69,047	168,712	1,448,494	Tokenis
2. History	2,119,048	93,721	95,312	2,308,081	1
3. Literature	1,753,839	205,482	26,616	1,985,937	Lemma
4. Linguistics	1,544,166	156,921	228,058	1,929,145	Dector
5. Philosophy	454,266	48,157	140,892	643,315	Pos-tag
6. Religious studies	2,282,125	48,794	288,615	2 619,534	
Total (tokens)	9,364,179	622,122	948,205	10,934,506	

Freely accessible through The Swedish Language Bank

Att ära Jesus (innebär) att älska sin nästa som sig själv. Denna kombination innebär en modifiering av den tidigare grundmodellen, se figur 1 avsnitt 4.2.2. För individen innebär privatisering att de funktioner som ger mening och tillhörighet är styrda av r er mening och tillhörighet är styrda av relationer i den privata sfären, vilket innebär att individens källa till identitet huvudsakligen finns inom den privata sfärer Detta innebär inte att predikan ses som en företeelse utan förankring ett konkret samfu n från början sätts den enskilde pastorn in i ett specifikt sammanhang som innebär att hon / han ser på sig själv som en predikant och förkunnare. andra " som Anna talar om, tillför inget till förståelsen av vad nästankärlek innebär Individualisering innebär att individen och hennes / hans behov blir mer tydligt urskilda från den används i det här sammanhanget tillägget institutionell differentiering, vilket innebär att olika institutioner har skiljts från varandra, inte bara stat och religio 29 Individualisering får konsekvenser för individens religiositet och innebär att det inte längre finns ett samhälligt stöd för vare sig religiösa övert nlingen som kollektiv ska skapa regler, medan kollektivistisk undersåtlighet innebär att församlingen inte ska skapa regler Denna solidariska hållning innebär ett högt värderande av att följa regler och ett högt värderande av att skapa reg Att " Ge Kejsaren det honom tillhör ", måste väl innebära att ge också det som övrigt krävs för vårt gemensamma liv? erentiering och privatisering får konsekvenser för enskilda människor som **innebär** att olika livsvärldar hålls isär Detta missionsförbundiska värdesystem innebär att sociologiska bakgrundsfaktorer som i andra sammanhang ofta leder till skillnader i uppfattr Tyvärr innebar inte denna utvidgade förfrågan någon större skillnad. För Aagedal innebär privatiseringsteorin att förkunnelsen i första hand vänder sig till människornas privata sfär och

Web-interface and free downloadable list

• Compilation of corpora from other disciplines, e.g. social sciences, that can form the basis of a Swedish academic word list, which covers more university subjects

• Extraction of recurrent academic phrases from the corpora

• User studies, e.g. in relation to *The Language Support* service at the University of Gothenburg

• A Nordic collaboration with focus on language in academic settings in Denmark, Norway and Sweden

The work has partly been financed by the University of Gothenburg's Language Year