Providing comprehensible input where comprehensible input is unreachable

- teaching polysynthetic greenlandic L2

Per Langgård, head advisor, Oqaasileriffik

Abstract

Greenlandic is notorious among the world's languages for morphological complexity and by many - scholars and laymen alike - held to be next to unacquirable.

Before the new constitution of 1953 the few Danish officials in Greenland obtained some Greenlandic pidgin after years in Greenland and a few missionaries and trade managers actually became fluent speakers. After 1953 less than one per mil out of many thousands of Danes who went to work in Greenland and enrolled in language classes made it to functional Greenlandic competence.

In the 80-s Greenlandic like almost all other L2 classrooms was heavily influenced by inductive didactics (ELT) in an attempt to improve results but results were poor. Even worse than known from traditional deductive teaching.

With language technology supported "Greenlandic for Foreigners" (2009 and 2011) introducing 'integrated focus on forms' didactics things improved dramatically. Still, drop-out rates were much too high in spite of obvious improvements. Students seldom learn to acquire Greenlandic via natural communication.

Krashen's controversial old comprehensible input theory no doubt encompasses considerable explanatory power as well as a usefull roadmap for improving the situation: Greenlandic has an extremely low type-token relation, derivational powers building very long complex words, about four phonological changes per running word, complicated inderivational syntax etc. To put it short: The simplest of sentences is complicated far beyond a student's capacity for a long time and therefore not exploitable for acquisition.

Automated semantically tagged word and sentence generators efficiently deliver randomized comprehensible input targeted at learners' actual lexical and grammatical competence at any given time. Existing automata and machine generated perception and production exercises for Greenlandic will be introduced together with work in progress for next generation apps.

A bit of background

History ultra short

Greenland was colonized in 1721. It became an integrated part of Denmark with the new constitution of 1953 which gave rise to a massive modernization of the

Greenlandic society and put quite some pressure on the culture and language in Greenland. But already from the late 60-s critical voices against Danish imperialism were more and more often heard until Greenland ultimately gained Home Rule in 1979. In 2009 Home Rule was expanded into Self Government. A public and political wish for independency is widespread but a solid majority in parliament (*Inatsisartut*) calls for postponement of the decision to better prepare the Greenlandic society for the challenges that come with independency.

Greenlandic

Greenlandic or Kalaallisut is the biggest and most vital dialect in the esk-aleut family of languages. It is typologically polysynthetic. Before 1953 Greenland was monolingual in most respects including school and church with the local language in high esteem. By the early 19th century the Bible was translated into Greenlandic, illiteracy basically done with, several high-class grammars and dictionaries had been produced and in 1851 the first standard orthography was decided. It lasted until 1973 when a new phonomatic orthography was introduced.

Greenlandic alway was **the** language in Greenland and with Home Rule in 1979 its status as Greenland's primary language was included in the Home Rule Act and with the Self Government Act of 2009 Greenlandic constitutionally became Greenland's only official language.

In spite of the pressure from globalization on all small languages Greenlandic is still extremely vital considering that there only are about 57,000 inhabitants in Greenland. Language is not recorded in the national registers so no one knows for sure how many speakers of Greenlandic there are, but at *Oqaasileriffik* (the government's official bureau for national language questions) a fair guess says that about 50% of the population are monolinguals in Greenlandic with no or limited command of Danish L2, 25% are bilinguals more or less balanced and 25% are monolingual in Danish with no or extremely limited command of Greenlandic L2.

A brief introduction to polysynthesis

Greenlandic is highly polysynthetic which means that lots of (most of?) semantic and grammatical information that in analytic languages like English or Danish - and to a lesser degree also in synthetic languages like Finnish – are expressed with individual adjectives, pronouns, modal verbs, adverbs etc. in Greenlandic is integrated or inderived into complex words that often grow very long. Here is an illustrative example:

oqaaseq is a noun meaning *a word*
(noun+)*LIRI* is a derivational morpheme meaning *to deal with* (noun)
-voq is an inflectional morpheme covering 'indicative mood in 3Sg'
(verb stem+)VIK is a derivational morpheme meaning *a place where one* (verb)
(verb stem+)NIQ is a derivational morpheme meaning *the act of* (verb)[1]
Put these pieces together obeying sandhi-rules and phonology and get for a start

oqaasilerivoq – he deals with words or *he works with linguistics*
Oqaasileriffik – the place where they deal with words or *The Language Secretariat*
oqaasilerineq – the act of dealing with words or *linguistics*

1 Speakers of Finnish will see the parallel to *-minen*

Add a few more morphemes from the stock of hundreds to your mental lexicon like

(noun+)PAK is a derivational morpheme meaning *a number of* (noun)

(noun+)SUAQ is a derivational morpheme meaning *a big* (noun)

and get new words like

(adding PAK) *oqaaserpaat – a number of words* and *Oqaasileriffippaat – a number of Language Secretariats*[2]

(adding SUAQ) *oqaasersuaq – a big word* and *Oqaasileriffissuaq – the great Oqaasileriffik*

You can also combine an in principle unlimited number of morphemes

(combine PAK and SUAQ) *oqaaserpassuit – a lot of words* and *Oqaasileriffippassuit – a lot of Language Secretariats*[3]

(combine SUAQ and PAK) *oqaasersuarpaat – a number of big words* and *Oqaasileriffissuarpaat – a number of great Oqaasileriffiks*

From here you can go on constructing words almost unlimitedly including

oqaaserpassualerineq – language technology

oqaaserpassualerivoq – he works with language technology

Oqaaserpassualeriffik – the department of language technology

Now add a few extra morphemes like

(verb+)LIR meaning *begin* (verb)

(verb+)QQAAR meaning (verb) for the first time

(noun+)QAR meaning *have/there is* (inderived noun)

-mat is inflection covering 'causative mood in 3Sg'

and you can form a large number of good useable words like

(*oqaaseq*+LIRI+VIK+QAR+LIR+QQAAR+{mat}) *1999-imi Oqaasileriffeqaleqqaarmat .. / In 1999 when Oqaasileriffik for the first time started to exist ..*

(*oqaaseq*+PAK+SUAQ+LIRI+VIK+QAR+LIR+QQAAR+{vuq}) *2005-imi Oqaaserpassualeriffeqaleqqaarmat .. / In 2005 when the department of language technology for the first time started to exist ..*

In Greenlandic word formation is a highly creative proces where morphemes combine rather freely but not entirely freely

(*oqaaseq*+QQAAR+LIRI+{vuq}) **oqaaseqqaalerivoq* and (*oqaaseq*+LIRI+QAR+{vuq}) **oqaasilereqarpoq*

and an unlimited amount of others like these are **not** correct words.

It should also be considered how overwhelming the combinatoric powers of a polysynthetic language are. For example will the 81 stems, 12 derivational morphemes, 15 endings and 2 enclitics introduced by the end of the first chapter in *Greenlandic for Foreigners. Module II* (2011) generate far above 2 million morphologically correct wordforms , but a large proportion of these millions of words will be semantically unusable words like **Oqaasileriffippaat* mentioned in footnote 2. This fact should be considered carefully since it is of paramount importance to the discussion below about input versus production as pedagogical strategies in the non-analytic language L2 classroom.

Greenlandic L2

Whereas Greenlandic L1 can be considered a clear-cut success in comparison with all other small languages in the world the same obviously cannot be said about Greenlandic L2. On the contrary. In spite of the fact that the Greenlandic politicians

2 A highly unlikely word since *Oqaasileriffik* is a unique proper noun. Accordingly, **Oqaasileriffippaat* is a morphologically correct but unusable word

3 See footnote 2 above

strongly advocate that Danes and crossbreeds should learn and speak Greenlandic and that most non-Greenlandic speakers have a wish to acquire Greenlandic almost none get even close to functional competence in Greenlandic.

Of the thousands of learners who entered the Greenlandic L2 classrooms over the last 50 years only a handful ever made it to function on the Greenlandic speaking labour market even at a level of competence far below near-nativeness. It is most likely a fair guess that less than one per mil of all learners acquired Greenlandic enough to teach in school or at university even if accepting strongly broken Greenlandic as a qualifying instruction language.

In earlier days Greenlandic L2 was taught in the latin tradition with heavy emphasis on grammar and translation skills. (Bugge:1952) and (Bugge:1966) are illustrative of the epoch. A handful of fairly qualified L2 speakers came out of that tradition in the 50-s and 60-s.

In 1973 Keld Thor Pedersen produced a rather comprehensive system, *Grønlandsk for begyndere I og II,* based on audio-lingual methodology. It meant a great leap forward.[4] Unfortunately, the system was given up too soon primarily because the author did not want to adapt/renew the system to the new standard orthography of 1973. A handful of successful learners from the 70-s including myself got their introduction to Greenlandic through Keld Thor Pedersen's system.

Concomitant with the shift in pedagogical thinking from linguistic competence toward communicative competence especially in English L2 teaching Home Rule government had new systems for Greenlandic L2 developed along these lines in the 80-s and 90-s and a number of L2 specialists were imported both as tutors for the teachers of Greenlandic L2 and to teach at Home Rule's L2 academy established in 1997.

Unfortunately, none of the specialists spoke Greenlandic and neither they nor the Greenlandic Government remembered to ask whether one a priori should trust that didactics that work well when teaching analytic English L2 automatically will work when teaching polysynthetic Greenlandic.

ELT didactics have not worked well! I ahve in a number of publications (in Danish e.g. Langgård:2014) argued that attempts to base Greenlandic L2 on ELT methodology (not least 'situational approach') with grammar playing an insignificant role has been a big mistake. I have no doubt that Greenlandic L2 will stay deadlocked until we get rid of the ELT harnish and get going with developing methodology for Greenlandic L2 based on the language itself developed by applied linguists and language teachers who know the language themselves.

The basic problem in a nutshell

As demonstrated by the counting above 110 individual pieces of language material will generate millions of wordforms. For me it goes without saying that acquisition cannot happen neither by accumulating and internalizing individual words and phrases as in the ELT classroom nor via paradigmatic studies combined with dictionary studies the latin way. Acquiring a polysynthetic language like Greenlandic

4 The pedagogical and methodological bases for my *Learn Greenlandic* are clearly inspired from *Grønlandsk for begyndere*.

is much more a matter of mastering the word formation process than accumulation of individual words.

Unfortunately, there are no quick fixes to obtain the capacity to process Greenlandic in the unconscious way that is needed. The verbal paradigms alone cover more than 300 individual inflectional morphemes. Add hereto one of the most complex derivational systems in the world. I have calculated that derivational and inflectional morphemes combine at an average rate of about 4 times per running word every time changing the "stem" to the left through deletion, assimilation or replacement. No one will internalize such masses of morphemes and processes to be useable fast enough for normal everyday communication through theoretical studies.

The achievement might seem overwhelming but obviously is doable. Almost all Greenlanders know how to do it, Danish children immigrating to Greenland at an early age get to do it and a handful of individuals who were introduced to Greenlandic as adults learned how to do it.

After almost 50 years of theoretical studies in applied linguistics and after trial-and-error encounters with thousands of learners over the years I feel comparatively safe in pinning out a few headlines about dos and don'ts in the Greenlandic L2 classroom[5]

- One does **not** acquire a morphologically rich language[6] from primers, dictionaries, teachers, compendia or the like. Such are shortcuts for you to outline the targets you are about to acquire. Acquisition is an individual process as is the magical transition of passive grammar to the active (productive) grammar inside the learner's head.
- Neglecting grammar is both useless and strongly demotivating to most adult students. Grammar is unavoidable but it should form an active mental grammar not a passive lookup system.
- It is a bad idea to wet your pants to stay warm in the Arctic! Teaching phrases and situations gives the learners an immediate satisfaction but does not pave the way to internalized acquisition of a polysynthetic language. On the contrary will such strategies produce lexicons of fossilized knowledge that eventually block real acquisition.
- With the rich morphology in the language words and sentences can combine almost unlimitedly. In theory! There are lots of usus-rules and semantic constraints reducing freedom to form utterances and a number of sandhi-rules to consider when generating Greenlandic language. It is therefore a very bad idea to let translation exercises from mother tongue to target language prevail in teaching stategies just as is urging students to speak the target language off the top of their heads in the early stages. Native Greenlandic **must** be both source and target in the proces. It is therefore mandatory to base all activities ab initio up till at least the first few hundreds hours of struggling with the language on reception (*comprehensible input*)**,** not on production. Attempts to produce L2 too early **will** backfire.

---

5 From personal experience as a student of Finnish L2 I am convinced that the statements are valid also with synthetic languages with a lower level of synthesis as Finnish

6 Whether this statement is true or false when teaching analytic languages is beyond my experiences. My recent experiences stem from Greenlandic and Finnish L2

Integrated focus on forms

I maintained above that deductive pedagogy has a role to play in the Greenlandic L2 classroom. As demonstrated by a number of scholars in recent years including Spada & Lightbown (see Lessard-Clouston:2011) many situations and environments call for lesser emphasis on induction and more on deduction. Such factors include considerable typological distance between mother tongue and target language, morphological complexity in the target language, learners' age and academic background[7] and several other parameters.

But it should be stressed that deduction here is used very differently from grammar training in the latin tradition. Paradigmatic approaches are not advised. The kernel of class-room activities must be input of authentic target language with explanations and explanatory pattern drills included only when needed.

The concept has been termed 'integrated focus on forms'. Here is one example:

In an input sentence in an early stage of learning a student who already learned that *nuliaq* means *wife* will meet the word *nuliara* and find out that the word means *my wife*. The student might jump to the wrong conclusion that the possessive pronoun, *my,* is an inflectional ending *ra and want to say for instance *illu/house*+ra instead of correct *illuga/my house.* In that situation the student should deductively be told that the inflectional morpheme adding a 1Sg possessor to a singular possessum is {ga} and that a phonological rule fuses any /q/ followed by /g/ to one and only one /r/. Now neither *nuliaq+ga = nuliara (my wife)* nor *illu+ga = illuga (my house)* should be problems and the next step in efficient teaching will be to make the student encounter as many {ga}-s as possible in authentic text.

Observe that whereas traditional grammar training would introduce the student to a paradigm of possessive inflection integrated focus on forms does not do that. In this framework deductive grammar will not be included before the student actually comes across problems in authentic language that call for explanations.

Input versus production – the comprehensible input theory

All successful students of Greenlandic I am acquainted with share certain steps in the acquisition process: They read or heard a word, understood it semantically, understood the morphemes that made up the word and extracted the rules at play. The next word and the next word again they processed in the same way probably shared some of the linguistic material or rules encountered in formerly processed words and sentences. Every word processed and understood this way accordingly added a fraction of help in deciphering other words. Here is an example:

Let *atuar = to read,* GUSUP is a derivational morpheme meaning *want to* (verb) and {punga} is an inflectional morpheme covering 'indicative mood + 1Sg' be new information. Combining the information into the word *atuarusuppunga/I want to read* includes among others the learner's ability to decipher the fusion of /r/ and /g/ for him to recognize the morpheme GUSUP. Now imagine that the student was given the

7 Grown-up students prefer more deduction than youngsters and children, and students having completed higher education more than students with less academic background

opportunity to hear/read and process a number of words with possessive {ga} mentioned above. He will then have achieved an understanding of the fusing process. One could say that he has obtained the ability to “see” the /g/, that disappeared from GUSUP in words like *atuar*+GUSUP = *atuarusup- (to want to read)* as well as the /g/ from {ga} in *nuliaq*+ga = *nuliara (my wife).*

This emphasis on authentic receptive language got a theoretical framework by the somehow controversial American scholar, Stephen Crashen, who termed the approach *The Comprehensible Input Theory* (see for instance Krashen:1988). According to the theory input of real language and a lot of it should be the predominant feature in the L2 classroom. Productive exercises and training should be limited and should be absolutely absent in the earlier phases of acquisition. Krashen terms the absence of productive activities in the beginning *The Silent Period*. According to Krashen the ability to produce L2 will appear in an almost magical way once the learner has been exposed to sufficient *comprehensible input.*

I am fully aware of the fact that Krashen’s theories not are unchallenged – far from. Nevertheless, Krashen’s ideas are basic to my thinking and my system because I never found an alternative that better than his works explained my students’ and my own problems.[8]

The paramount problem of *incomprehensible input*

However captivating *the comprehensible input theory* might be real life present all learners with an almost insurmountable problem. In a polysynthetic language even simple utterances inevitably will include morphemes that are unknown to a beginner. Very often the consequence is that one morpheme of unknown language added to stems and morphemes that could have been recognizable renders the whole construction incomprehensible and therefore adds nothing to the student’s perceptional capacity.

As a simple example of incomprehensibility consider a theoretical first welcome: *Perimik ateqarpunga. Qallunaajuvunga/ My name is Per. I am a Dane.*

To understand the three greenlandic words one will need to know that *ateq* means *name*, that QAR is a morpheme meaning *to have,* that *qallunaaq* means *Dane*, that U is a morpheme meaning *to be* and that *-vunga* and *-punga* both are endings with content ‘indicative mood 1Sg’. Getting such pieces of knowledge into one’s mental lexicon seems not to be a bigger problem than learning the 7 words in English had English been the target language.

But one will furthermore need to know that (i) *-vunga* is used after vowels and *-punga* after consonants, (ii) that arguments to inderived object like *Per* (an argument to *ateq/name* in the verbs *ateqar-/ to_have_name*) will be in the instrumental case (-

8 I have even applied the theories to myself in practice. In the spring of 1990 I needed to learn Finnish in Nuuk, Greenland. There were no speakers of Finnish around so I sat down about an hour a day with tapes and text for Olli Nuutinen’s *Suomea suomeksi* and Eila Hämäläinen’s *Suomen harjoituksia.* I stuck painstakingly to Krashen’s theories focusing on perception, understanding not only semantics but also structure and reproduced – not produced!! - what I understood. *The silent period* was thus absolute for me. I refrained from producing any Finnish at all. Still, when I made it to Finland in June Krashen’s magic actually happened. I was able to somehow speak broken Finnish and understood enough of spoken words, road signs etc. to actually make my way in many situations

mik), (iii) that both QAR/*to have* and U/*to be* truncate final consonants and that (iv) two /a/-s cannot be followed by a third vowel so that j-epenthesis is mandatory.

It goes without saying that to an L2 beginner problems deciphering the Greenlandic input is not even comparable with the problems a student of English will have to face. The English input is comprehensible almost immediately the Greenlandic input is incomprehensible for a long time.

I do not doubt that the unachievability of comprehensible input is the crucial single factor that explains why so few learners ever acquire Greenlandic. Without access to comprehensible input the process of internalizing the first bits and pieces that will help you to understand the next bits and pieces needed to understand the language will not start to evolve.

How to provide comprehensible input when everything in the real world is too complex to be comprehensible

Greenlandic language technology has reached a state of refinement where impossibility of providing comprehensible input no longer is an unsurmountable issue. Here is what we recently started to do:

After the first chapter in the primer a finite state automaton covering all lexical and grammatical material presented in the primer is build. One could say that we created a small but full language with a few proper nouns inflected in a few cases. In casu the first lesson comprises 17 place names and three case endings plus a few pieces of information about the two declension classes needed to cover the place names presented. Finally two general phonological rules are presented.

With these pieces of information the learner has tools enough to decipher all of the 51 wordforms this “lesson 1 language” comprises.

Next, the system will generate and randomize the wordforms and present them to the student with different assignments presented both as written text and spoken out using speech synthesis.

The student thus has the opportunity to get going with gradually building an understanding for sandhi and sound changes and to start the personal travel toward a state where sound changes etcetera are so integrated that they finally will be accepted/ understood unconsciously thus facilitating the next steps. Future unknown pieces of language will much easier be isolated when perception no longer is tampered by difficulties unerstanding already known parts of the language.

In chapter 2 14 proper nouns, 4 verbal stems, 1 pronoun and 5 inflectional morphemes are introduced. When added to “lesson 1 language” the new “lesson 2 language” comprises 136 wordforms that can be randomized in the same way as in chapter 1. And with the introduction of a few verbs it is now possible to randomize hundreds of small sentences containing nothing but comprehensible input:

Nuka Upernavimmit aallarpoq:
(Nuka from_Upernavik he_left) *Nuka left from Upernavik*

Aputsiaq Sisimiunut tikippoq:

(Aputsiaq to_Sisimiut he_arrived) *Aputsiaq came to Sisimiut*

Pituffimmit tikippunga:
(from_Pituffik I_arrived) *I came from Pituffik*

Rosing Tasiilami ilinniartippara:
(Rosing in_Tasiilaq I_taught_him) *I taught Rosing in Tasiilaq*

Challenges with the sentence generator

The challenges contained in Chomsky's famous quote *Colorless green ideas sleep furiously* of course must be considered when creating a sentence generator. In the onset the generator will happily generate morphologically well-formed but semantically unusable sentences like **Nuka Upernavimmi aallarpoq/ Nuka left in Upernavik* or **Rosing Tasiilamut ilinniartippara/I taught Rosing to Tasiilaq*.

Among other remedies to handle this problem all lexical entries are semantically tagged using Danish Framenet[9]. The tags enable developers to control how words combine during the automated generation process.

A small selection of rules from the syntax generator will illustrate the idea:

Invisible to the student *Rosing* is tagged Sem/Hum, *Tasiilaq* Sem/Geo and *ilinniartip-* Sem/teach. Among other tags could *Katuaq* (the Culture House in Nuuk) tagged with Sem/inst and *naapip-* (to meet (object)) tagged with Sem/encounter be mentioned.

One rule will state that any Sem/teach OR Sem/encounter will combine with Lok ('in-case') + Sem/geo OR Sem/inst but not with Trm ('to-case') OR Abl ('from-case')

Another rule will block anything but Sem/Hum as object for Sem/teach OR Sem/encounter

Accordingly an infinite number of semantically nonsensical sentences like

**Tasiilaq Rosingimi ilinniartippara/ I taught Tasiilaq on Rosing*

will be blocked whereas all combinations that are not blocked will be accepted and passed on to the exercises.

Welcome to an endless amount of input that is comprehensible at whatever step you've reached in your personal acquisition process

I am convinced that the possibilities language technology offers us the teachers and text book writers for the first time ever enable us to offer natural acquisition via comprehensible input and gradual internalizing of the rules and structures needed to understand authentic Greenlandic. I believe it fair to say that language technology in the morphologically rich languages L2 classroom will mean a revolution to the whole field and soon be the game changer that can pave the way for a dramatic increase in the number of outsiders reaching functional competence in any non-analytic L2.

There are many rationales for my optimism. Among them should be mentioned

9 An adaption of Fillmore's Berkeley FrameNet taxonomy by Dr. Eckhard Bick

- Developing L2 technology for synthetic and polysynthetic languages takes a firm understanding of the language both in theory and in practice whereas demands for technical know-how is limited because of easy access to existing infrastructure and language technology specialist. Technology supported L2 can thus easily be anchored locally with limited need of imported specialists.
- It is far more cost-efficient to develop teaching materials exploiting automata and synthetic speech than the alternative that include printed publications and studio recordings with human speakers
- Online instruction will reach even remote places with small populations without access to a sufficient number of competent teachers of L2
- Revisions of teaching material will not be delayed by capacity problems in the publishing business and only limitedly by lack of funding. Online teaching and exercises can be updated/corrected cheaply, continuously and immediately following any trial-and-error process.

And it works. Already in the early stages of Greenlandic L2 acquisition rather complex language like the following example picked randomly among thousands of machine generated sentences build solely by pieces known to the student at any time is executable in the system and thus ready as comprehensible input for the learner:

Eqqaamaviuk uanga aqaguagu Kalaallit Nunaanni najugaqassanngitsunga?/ Did you remember that I shall not live in Greenland the day after tomorrow?

Aap, eqqaamavara./ Yes, I remembered it.

References

Bugge, Aage. *Grønlandsk Rejseparlør.* Nyt Nordisk Forlag (1952)

Bugge, Aage. *Dansk-grønlandske stiløvelser.* Ministeriet for Grønland (1966)

Hinkel, Eli (ed.). *Handbook of Research in Second Language Teaching and Learning. II.* Routledge 2011

Hämäläinen, Eila. *Suomen harjoituksia.* Helsinki 1989

Krashen, Stephen. *Second Language Acquisition and Second Language Learning.* Prentice Hall 1988

Langgård, Per. *Hvorfor kan danskere i Grønland – stadigvæk – ikke grønlandsk?* IN: Pedersen, Birgit Kleist et al. (ed.). *Grønlandsk kultur- og samfundsforskning 2013-14.* Ilisimatusarfik/Forlaget Atuagkat 2014

Lessard-Clouston, Michael. *Focusing on form: A teacher's update on form-focused instruction.* IN: CALR Liguistic Journal 1. (2011)

Nuutinen, Olli. *Suomea suomeksi 1-2.* SKS (1977)

Pedersen, Keld Thor. *Grønlandsk for begyndere*. Nyt Nordisk Forlag (1973)