rijksuniversiteit groningen

On the measurement of morphosyntactic distances
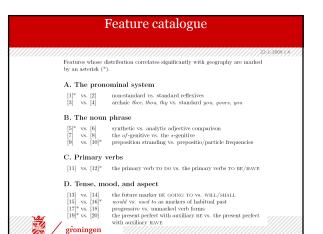
› Sebastian Kürschner

General problems

› The effect of morphosyntactic distances on intelligibility has – to our knowledge – never been systematically investigated

› Morphosyntactically driven dialectometry is still in the first steps

› How do you exclude differences on other levels than the morpho-syntactic one (e.g., phonetic differences) in experiments on intelligibility?

rijksuniversiteit groningen

Example: Research on morphosyntactic distances

› Szmrecsanyi (submitted) calculates morphosyntactic distances in English dialects

› frequency-based approach based on corpus analysis in FRED (Freiburg English Dialect Corpus)

› definition of 35 morphosyntactic variables

rijksuniversiteit groningen

Feature catalogue

Features whose distribution correlates significantly with geography are marked by an asterisk (*).

A. The pronominal system

| [1]* | vs. | [2] | non-standard vs. standard reflexives |
| [3] | vs. | [4] | archaic *thee, thou, thy* vs. standard *you, yours, you* |

B. The noun phrase

[5]*	vs.	[6]	synthetic vs. analytic adjective comparison
[7]	vs.	[8]	the *of*-genitive vs. the *s*-genitive
[9]	vs.	[10]*	preposition stranding vs. prepositio/particle frequencies

C. Primary verbs

| [11] | vs. | [12]* | the primary verb TO DO vs. the primary verbs TO BE/HAVE |

D. Tense, mood, and aspect

[13]	vs.	[14]	the future marker BE GOING TO vs. WILL/SHALL
[15]	vs.	[16]*	*would* vs. *used to* as markers of habitual past
[17]*	vs.	[18]	progressive vs. unmarked verb forms
[19]*	vs.	[20]	the present perfect with auxiliary BE vs. the present perfect with auxiliary HAVE

groningen

Feature catalogue

E. Verb morphology

[21]	vs.	[22]	*a*-prefixing on *-ing*-forms vs. bare *-ing*-forms
[23]	vs.	[24]	non-standard weak past tense and past participle forms vs. standard strong forms
[25]*	vs.	[26]	non-standard 'Bybee' verbs vs. corresponding standard forms
[27]			non-standard verbal *-s*
[28]*	vs.	[29]	non-standard past tense *done* vs. standard *did*
[30]	vs.	[31]	non-standard past tense *come* vs. standard *came*

F. Negation

[32]*	vs.	[33]	invariant *ain't* vs. *not/*n't/*nae*-negation
[34]*	vs.	[35]	multiple negation vs. simple negation
[36]*	vs.	[37]	negative contraction vs. auxiliary contraction
[38]*	vs.	[39]*	*don't* with 3rd person singular subjects vs. standard agreement
[40]	vs.	[41]	*never* as a preverbal past tense negator vs. standard negation

› etc.

rijksuniversiteit groningen

Szmrecsanyi (submitted)

› material in the corpus (varieties from 38 counties) was coded automatically and semi-automatically with the features

› feature frequency calculated and extracted into vectors

› table converted into distance matrix

rijksuniversiteit groningen

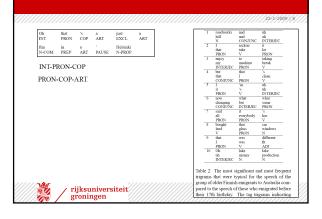## Example: Syntactic distances based on POS

› Nerbonne/Wiersma (2006)

› Analysis of a corpus tagged with POS
› analysis of 3-grams
› distance measure based on permutation tests

› corpus: English speech of two groups of Finnish emigrants to Australia:
  • emigrants older than 16 on the date of emigration
  • emigrants younger than 17 on the date of emigration

rijksuniversiteit groningen

Nerbonne/Wiersma

Oh	that	's	a	just	a
INT	PRON	COP	ART	EXCL	ART

fim	in	a	,	Helsinki
N-COM	PREP	ART	PAUSE	N-PROP

INT-PRON-COP

PRON-COP-ART

1	roadworks	and	uh
	hill	and	ah
	N	CONJUNC	INTERJEC
2	I	reckon	it
	that	take	lot
	PRON	V	PRON
3	enjoy	to	taking
	my	machine	break
	INTERJEC	PRON	V
4	but	that	's
	that	I	clean
	CONJUNC	PRON	V
5	I	'm	uh
	it	's	uh
	PRON	V	INTERJEC
6	now	what	what
	changing	but	some
	CONJUNC	INTERJEC	PRON
7	said	it	's
	all	everybody	has
	PRON	PRON	V
8	bought	that	car
	lead	glass	windows
	V	PRON	N
9	that	was	different
	I	was	fit
	PRON	V	ADJ
10	Oh	lake	lake
	uh	money	production
	INTERJEC	N	N

Table 2: The most significant and most frequent trigrams that were typical for the speech of the group of older Finnish emigrants to Australia compared to the speech of those who emigrated before their 17th birthday. The tag trigrams indicating

rijksuniversiteit groningen

Syntactic dialect atlas variables

› Spruit (2006)

› Syntactic distance measure based on 507 syntactic variables found in the Syntactic Atlas of the Dutch Dialects
  • variables and their variants were introduced by the editors of the atlas
  • frequency-based approach: occurrence of variables is aggregated using dialectometric methods

rijksuniversiteit groningen

Example variables

Table 1 Map 68a in SAND1 shows the five syntactic variables in the context of weak reflexive pronoun as object of inherently reflexive verb

Context:	Weak reflexive pronoun as object of inherent reflexive verb
Variables:	{zich, hem, zijn eigen, zichzelf, hemzelf}
Example:	Jan herinnert **zich** dat verhaal wel.
	Jan remembers himself that story AFFIRM
	"John certainly remembers that story."

Table 2 Map 82b in SAND1 shows the six syntactic variables in the context of short object relative

Context:	Short object relative
Variables:	{die, dat, wie, der, den/dem, as}
Example:	Dat is de man **die** ze geroepen hebben.
	That is the man who they called have
	"That is the man who they have called."

rijksuniversiteit groningen

Morphosyntactic distance and intelligibility

› all measures frequency-based
› how to model first-contact situation?
› how to test the influence of morphosyntactic distance?
› how to isolate the morphosyntactic level (or even morphological and syntactic levels separately) from other levels such as lexicon and phonetics?
› how to test the relative contribution of linguistic levels as determinants of intelligibility?

rijksuniversiteit groningen