

MULTILINGUALISM AND SOCIAL MEDIA

András Kornai
HAS Computer Science Research Institute
and
Department of Algebra, Budapest Institute of Technology

Vilnius

September 26 2013

WHAT IS SOCIAL MEDIA?

- Ostensive definition: Facebook, Twitter, Yelp, Foursquare, Youtube, ...
- Substantive definition: user-generated content, opportunity to discuss, ...
- Functional definition: social media is the new **proximity**
- Traditional proximity: by shared location, work, social fabric, ...
- What makes the new proximity different is that it is voluntary, not imposed by birth or tradition!
- We choose our FB friends, those we follow on Twitter, etc.
- Impact of traditional proximities remains very strong: we choose our high school buddies in FB, our colleagues on LinkedIn

WHAT IS SOCIAL MEDIA?

- Ostensive definition: Facebook, Twitter, Yelp, Foursquare, Youtube, ...
- Substantive definition: user-generated content, opportunity to discuss, ...
- Functional definition: social media is the new proximity
- Traditional proximity: by shared location, work, social fabric, ...
- What makes the new proximity different is that it is voluntary, not imposed by birth or tradition!
- We choose our FB friends, those we follow on Twitter, etc.
- Impact of traditional proximities remains very strong: we choose our high school buddies in FB, our colleagues on LinkedIn

WHAT IS SOCIAL MEDIA?

- Ostensive definition: Facebook, Twitter, Yelp, Foursquare, Youtube, ...
- Substantive definition: user-generated content, opportunity to discuss, ...
- Functional definition: social media is the new proximity
- Traditional proximity: by shared location, work, social fabric, ...
- What makes the new proximity different is that it is voluntary, not imposed by birth or tradition!
- We choose our FB friends, those we follow on Twitter, etc.
- Impact of traditional proximities remains very strong: we choose our high school buddies in FB, our colleagues on LinkedIn

WHAT IS SOCIAL MEDIA?

- Ostensive definition: Facebook, Twitter, Yelp, Foursquare, Youtube, ...
- Substantive definition: user-generated content, opportunity to discuss, ...
- Functional definition: social media is the new **proximity**
- Traditional proximity: by shared location, work, social fabric, ...
- What makes the new proximity different is that it is voluntary, not imposed by birth or tradition!
- We choose our FB friends, those we follow on Twitter, etc.
- Impact of traditional proximities remains very strong: we choose our high school buddies in FB, our colleagues on LinkedIn

WHAT IS SOCIAL MEDIA?

- Ostensive definition: Facebook, Twitter, Yelp, Foursquare, Youtube, ...
- Substantive definition: user-generated content, opportunity to discuss, ...
- Functional definition: social media is the new **proximity**
- Traditional proximity: by shared location, work, social fabric, ...
- What makes the new proximity different is that it is voluntary, not imposed by birth or tradition!
- We choose our FB friends, those we follow on Twitter, etc.
- Impact of traditional proximities remains very strong: we choose our high school buddies in FB, our colleagues on LinkedIn

WHAT IS SOCIAL MEDIA?

- Ostensive definition: Facebook, Twitter, Yelp, Foursquare, Youtube, ...
- Substantive definition: user-generated content, opportunity to discuss, ...
- Functional definition: social media is the new **proximity**
- Traditional proximity: by shared location, work, social fabric, ...
- What makes the new proximity different is that it is voluntary, not imposed by birth or tradition!
- We choose our FB friends, those we follow on Twitter, etc.
- Impact of traditional proximities remains very strong: we choose our high school buddies in FB, our colleagues on LinkedIn

WHAT IS SOCIAL MEDIA?

- Ostensive definition: Facebook, Twitter, Yelp, Foursquare, Youtube, ...
- Substantive definition: user-generated content, opportunity to discuss, ...
- Functional definition: social media is the new **proximity**
- Traditional proximity: by shared location, work, social fabric, ...
- What makes the new proximity different is that it is voluntary, not imposed by birth or tradition!
- We choose our FB friends, those we follow on Twitter, etc.
- Impact of traditional proximities remains very strong: we choose our high school buddies in FB, our colleagues on LinkedIn

WHAT IS SOCIAL MEDIA?

- Ostensive definition: Facebook, Twitter, Yelp, Foursquare, Youtube, ...
- Substantive definition: user-generated content, opportunity to discuss, ...
- Functional definition: social media is the new **proximity**
- Traditional proximity: by shared location, work, social fabric, ...
- What makes the new proximity different is that it is voluntary, not imposed by birth or tradition!
- We choose our FB friends, those we follow on Twitter, etc.
- Impact of traditional proximities remains very strong: we choose our high school buddies in FB, our colleagues on LinkedIn

LANGUAGE USE

- 'Heritage' languages (Latin, Manx) have no footprint in social media
- 'Still' languages (all but a few hundred of the world's languages) are digitally dead
- 'Vital' languages (almost all languages spoken at this meeting) have good uptake in social media
- 'Thriving' languages (the world's top 16-24) have excellent uptake
- The real frontier: bottom of 'vital' category
- Languages getting there: Basque, Welsh, Frisian, ...

LANGUAGE USE

- 'Heritage' languages (Latin, Manx) have no footprint in social media
- 'Still' languages (all but a few hundred of the world's languages) are digitally dead
- 'Vital' languages (almost all languages spoken at this meeting) have good uptake in social media
- 'Thriving' languages (the world's top 16-24) have excellent uptake
- The real frontier: bottom of 'vital' category
- Languages getting there: Basque, Welsh, Frisian, ...

LANGUAGE USE

- 'Heritage' languages (Latin, Manx) have no footprint in social media
- 'Still' languages (all but a few hundred of the world's languages) are digitally dead
- 'Vital' languages (almost all languages spoken at this meeting) have good uptake in social media
- 'Thriving' languages (the world's top 16-24) have excellent uptake
- The real frontier: bottom of 'vital' category
- Languages getting there: Basque, Welsh, Frisian, ...

LANGUAGE USE

- 'Heritage' languages (Latin, Manx) have no footprint in social media
- 'Still' languages (all but a few hundred of the world's languages) are digitally dead
- 'Vital' languages (almost all languages spoken at this meeting) have good uptake in social media
- 'Thriving' languages (the world's top 16-24) have excellent uptake
- The real frontier: bottom of 'vital' category
- Languages getting there: Basque, Welsh, Frisian, ...

LANGUAGE USE

- 'Heritage' languages (Latin, Manx) have no footprint in social media
- 'Still' languages (all but a few hundred of the world's languages) are digitally dead
- 'Vital' languages (almost all languages spoken at this meeting) have good uptake in social media
- 'Thriving' languages (the world's top 16-24) have excellent uptake
- The real frontier: bottom of 'vital' category
- Languages getting there: Basque, Welsh, Frisian, ...

LANGUAGE USE

- 'Heritage' languages (Latin, Manx) have no footprint in social media
- 'Still' languages (all but a few hundred of the world's languages) are digitally dead
- 'Vital' languages (almost all languages spoken at this meeting) have good uptake in social media
- 'Thriving' languages (the world's top 16-24) have excellent uptake
- The real frontier: bottom of 'vital' category
- Languages getting there: Basque, Welsh, Frisian, ...

LANGUAGE USE

- 'Heritage' languages (Latin, Manx) have no footprint in social media
- 'Still' languages (all but a few hundred of the world's languages) are digitally dead
- 'Vital' languages (almost all languages spoken at this meeting) have good uptake in social media
- 'Thriving' languages (the world's top 16-24) have excellent uptake
- The real frontier: bottom of 'vital' category
- Languages getting there: Basque, Welsh, Frisian, ...

TWEETS (AND BLOGS) IN VITAL LANGUAGES

Language	Users	Tweets	Top User	Tweets	First Tweet
Euskara	17042	4771516	berria	44318	eastigarraga
Kreyòl Ayisyen	14267	3859311	RoodHT509	49218	tichrist
Cymraeg	14218	2524678	newyddcymraeg	58651	meigwilym
Kapampangan	1379	1611553	itsmeshellie	19450	desperada
Frysk	2048	578132	omropfytsban	61705	vogelzang
Setswana	314	571908	sesutho	30945	WameDre
Gaelge	7389	489734	aonghusoha	25168	imeall
Asturianu	761	269808	lyanc	16717	Pingarates
Hausa	300	239591	bbchausa	20638	Richee_Rich
Soomaaliga	558	144348	Weedhsan	15987	HaPpYMaXaMeD
Gàidhlig	652	95555	sconewt	27303	ruighean
Māori	339	90425	maonewt	29810	DigitalMaori
Gaelg	26	85760	glvoid	40479	greinneyder
Chamoru	9	61916	chanewtest	25699	jenccamacho
Tamasheq	3	57363	tmjnew	27484	tmjbbi
Aragonés	80	55345	Purnas	13716	jesmar
عربي	25	51215	mvblogs	9788	muraasil
Winaray	33	41208	llovetacloban	6750	plamaamendz
Malagasy	252	37233	Ainjah	4514	elsifaka
Chicheŵa	143	33714	skinnybrizzo	6866	caseyeshbit
יידיש	260	31143	jewish_wire	11553	ZackBergerMDPhD
Cántabru	14	30784	DieguSG	6043	r_macho
Occitan	98	29645	occitania	8736	sbagudanch
Emiliàn e Rumagnòl	53	28406	ingiro_concerti	2349	Ze Beeps

INDIGENOUS TWEETS.COM

Blog

Indigenous Blogs

Kevin Scannell

POLICY SUGGESTIONS

John Wyles asked for “practical suggestions for policymakers on how to define their priorities in using social media to maximise the impact of their communications and information policies”.

- Work the chain of literacy
- Think outside the social media box
- Make a big push for transparency

POLICY SUGGESTIONS

John Wyles asked for “practical suggestions for policymakers on how to define their priorities in using social media to maximise the impact of their communications and information policies”.

- Work the chain of literacy
- Think outside the social media box
- Make a big push for transparency

POLICY SUGGESTIONS

John Wyles asked for “practical suggestions for policymakers on how to define their priorities in using social media to maximise the impact of their communications and information policies”.

- Work the chain of literacy
- Think outside the social media box
- Make a big push for transparency

POLICY SUGGESTIONS

John Wyles asked for “practical suggestions for policymakers on how to define their priorities in using social media to maximise the impact of their communications and information policies”.

- Work the chain of literacy
- Think outside the social media box
- Make a big push for transparency

THE CHAIN OF LITERACY IN THE DIGITAL AGE

- Standardised orthography, dictionaries (incl. monolingual!)
- Word-level tools: spellchecker, stemmer, morphological analysis
- Phrase-level tools: part-of-speech tagging, named entity recognition
- Sentence-level tools: light parsing, machine translation
- Input: speech and character recognition
- Output: generation, synthesis, fonts

THE CHAIN OF LITERACY IN THE DIGITAL AGE

- Standardised orthography, dictionaries (incl. monolingual!)
- Word-level tools: spellchecker, stemmer, morphological analysis
- Phrase-level tools: part-of-speech tagging, named entity recognition
- Sentence-level tools: light parsing, machine translation
- Input: speech and character recognition
- Output: generation, synthesis, fonts

THE CHAIN OF LITERACY IN THE DIGITAL AGE

- Standardised orthography, dictionaries (incl. monolingual!)
- Word-level tools: spellchecker, stemmer, morphological analysis
- Phrase-level tools: part-of-speech tagging, named entity recognition
- Sentence-level tools: light parsing, machine translation
- Input: speech and character recognition
- Output: generation, synthesis, fonts

THE CHAIN OF LITERACY IN THE DIGITAL AGE

- Standardised orthography, dictionaries (incl. monolingual!)
- Word-level tools: spellchecker, stemmer, morphological analysis
- Phrase-level tools: part-of-speech tagging, named entity recognition
- Sentence-level tools: light parsing, machine translation
- Input: speech and character recognition
- Output: generation, synthesis, fonts

THE CHAIN OF LITERACY IN THE DIGITAL AGE

- Standardised orthography, dictionaries (incl. monolingual!)
- Word-level tools: spellchecker, stemmer, morphological analysis
- Phrase-level tools: part-of-speech tagging, named entity recognition
- Sentence-level tools: light parsing, machine translation
- Input: speech and character recognition
- Output: generation, synthesis, fonts

THE CHAIN OF LITERACY IN THE DIGITAL AGE

- Standardised orthography, dictionaries (incl. monolingual!)
- Word-level tools: spellchecker, stemmer, morphological analysis
- Phrase-level tools: part-of-speech tagging, named entity recognition
- Sentence-level tools: light parsing, machine translation
- Input: speech and character recognition
- Output: generation, synthesis, fonts

THE CHAIN OF LITERACY IN THE DIGITAL AGE

- Standardised orthography, dictionaries (incl. monolingual!)
- Word-level tools: spellchecker, stemmer, morphological analysis
- Phrase-level tools: part-of-speech tagging, named entity recognition
- Sentence-level tools: light parsing, machine translation
- Input: speech and character recognition
- Output: generation, synthesis, fonts

OUTSIDE THE SOCIAL MEDIA BOX

- It takes PhD-level people to develop the digital language tools
- Must support higher education of native speakers
- Look outside EU: India, much of Africa and Asia struggle with the same problems
- Look outside social media proper: email, hangouts, voice calls (not just Skype), text/sms all thrive when the tools are there

OUTSIDE THE SOCIAL MEDIA BOX

- It takes PhD-level people to develop the digital language tools
- Must support higher education of native speakers
- Look outside EU: India, much of Africa and Asia struggle with the same problems
- Look outside social media proper: email, hangouts, voice calls (not just Skype), text/sms all thrive when the tools are there

OUTSIDE THE SOCIAL MEDIA BOX

- It takes PhD-level people to develop the digital language tools
- Must support higher education of native speakers
- Look outside EU: India, much of Africa and Asia struggle with the same problems
- Look outside social media proper: email, hangouts, voice calls (not just Skype), text/sms all thrive when the tools are there

OUTSIDE THE SOCIAL MEDIA BOX

- It takes PhD-level people to develop the digital language tools
- Must support higher education of native speakers
- Look outside EU: India, much of Africa and Asia struggle with the same problems
- Look outside social media proper: email, hangouts, voice calls (not just Skype), text/sms all thrive when the tools are there

OUTSIDE THE SOCIAL MEDIA BOX

- It takes PhD-level people to develop the digital language tools
- Must support higher education of native speakers
- Look outside EU: India, much of Africa and Asia struggle with the same problems
- Look outside social media proper: email, hangouts, voice calls (not just Skype), text/sms all thrive when the tools are there

TRANSPARENCY

- Corpora, the lifeblood of modern computational linguistics, must be unencumbered by copyright. **Make sure a research exemption is enshrined in the legal framework**
- National projects need to make their corpora not just searchable but also downloadable by ROAMing (randomize, omit, anonymize, mix). **Make sure open access to materials collected is a precondition of EU funding**
- A wikipedia is a necessary but insufficient condition for digital ascent, digital communities (not just read-only material) are also needed. **Give micro-grants to small communities (literary, theatrical, whatever brings people together) to document in their native language what they are doing.**

TRANSPARENCY

- Corpora, the lifeblood of modern computational linguistics, must be unencumbered by copyright. **Make sure a research exemption is enshrined in the legal framework**
- National projects need to make their corpora not just searchable but also downloadable by ROAMing (randomize, omit, anonymize, mix). **Make sure open access to materials collected is a precondition of EU funding**
- A wikipedia is a necessary but insufficient condition for digital ascent, digital communities (not just read-only material) are also needed. **Give micro-grants to small communities (literary, theatrical, whatever brings people together) to document in their native language what they are doing.**

TRANSPARENCY

- Corpora, the lifeblood of modern computational linguistics, must be unencumbered by copyright. **Make sure a research exemption is enshrined in the legal framework**
- National projects need to make their corpora not just searchable but also downloadable by ROAMing (randomize, omit, anonymize, mix). **Make sure open access to materials collected is a precondition of EU funding**
- A wikipedia is a necessary but insufficient condition for digital ascent, digital communities (not just read-only material) are also needed. **Give micro-grants to small communities (literary, theatrical, whatever brings people together) to document in their native language what they are doing.**

TRANSPARENCY

- Corpora, the lifeblood of modern computational linguistics, must be unencumbered by copyright. **Make sure a research exemption is enshrined in the legal framework**
- National projects need to make their corpora not just searchable but also downloadable by ROAMing (randomize, omit, anonymize, mix). **Make sure open access to materials collected is a precondition of EU funding**
- A wikipedia is a necessary but insufficient condition for digital ascent, digital communities (not just read-only material) are also needed. **Give micro-grants to small communities (literary, theatrical, whatever brings people together) to document in their native language what they are doing.**

CONCLUSIONS

- It is not just your social media content that large-scale aggregators have visibility into, the NSA, your ISP, and your software vendors are already analyzing your phonecalls/text (and perhaps your hard drive as well). **Make sure their information monopoly is broken**
- Change starts with you. If you are a public servant, especially a decision-maker disbursing public funds, **make your own social network transparent**. It's not the NSA you want to see it, it's your constituents.

CONCLUSIONS

- It is not just your social media content that large-scale aggregators have visibility into, the NSA, your ISP, and your software vendors are already analyzing your phonecalls/text (and perhaps your hard drive as well). **Make sure their information monopoly is broken**
- Change starts with you. If you are a public servant, especially a decision-maker disbursing public funds, **make your own social network transparent**. It's not the NSA you want to see it, it's your constituents.

CONCLUSIONS

- It is not just your social media content that large-scale aggregators have visibility into, the NSA, your ISP, and your software vendors are already analyzing your phonecalls/text (and perhaps your hard drive as well). **Make sure their information monopoly is broken**
- Change starts with you. If you are a public servant, especially a decision-maker disbursing public funds, **make your own social network transparent**. It's not the NSA you want to see it, it's your constituents.