KYO KAGEURA

KEITA TSUJI

National Institute of Informatics

Tokyo, Japan

{kyo,keita}@nii.ac.jp

On the nature of "Japanese language" in news reporting

1. Introduction

Individual languages that we talk about in linguistics and related areas such as computational linguistics —Japanese, French, Chinese, English, Spanish, etc.— are social constructs/institutions and not natural objects (Saussure, 1910/11 [1993]). It is now well established that print media contributed significantly to the consolidation of "individual languages" (and "nationalism"/"nation-state") as we talk of now (Anderson, 1991; Febvre & Martin, 1971), which Saussure also seemed to be well aware of (Karatani, 1995). If individual languages that we talk of now is priviledgedly conditioned on the print media, then it is not unnatural to ask: does the internet revolution of communication affect languages and may contribute to different arrangement of "individual languages", which from the point of view of individual languages may mean the disollusion of unity of individual languages that is currently taken for granted? Our long term aim is to establish an epistemological and methodological framework within which this problem can be properly addressed; our short term aim is to observe some factors or concrete changes that might contribute to this (possible) radical dissolusion of individual languages within the current use of languages in different communication media. In this paper, we formulate the problems and report the result of a very preliminary analysis we have carried out about what is currently called "Japanese".

2. Language and the Internet — Existing Studies

Growing number of studies are devoted to the changes in a language introduced by electronic means of communication (e.g. Aitchison & Lewis, 2003; Baron, 1998; Crystal, 2001; Ferrara et al., 1991). For instance, Baron (1998), Crystal (2001) and Ferrara et al. (1991) dealt with, among others, the topic concerning whether language/discourse used in e-mails are more like written language/discourse or spolen language/discourse. Crystal (2001) examined the languages of e-mail, of chatgroups, of virtual worlds and of the Web from the point of view of their characteristics vis-a-vis conventional features observed in written and spolen languages/discourse.

Most of them, however, assume the stablity of "individual languages" such as English or Spanish or Japanese as we currently talk of, and observe changes within a language, e.g. new usage of existing linguistic items, emergence of new lexical items, new discourse types, new register, etc etc. Of course, given the lack of theoretico-conceptual devices that enables us to talk about languages without referring to "individual languages" that we talk of (Saussure, from Constantin's note, seemed to try to consolidate the concept of "la langue" independently of what we currently talk of under the term "individual language", but he ended up abandoning talking about language completely). What is lacking in the existing studies of new media languages is the path from new discourse/communication not only to the changes in "individual languages" but changes to the very framework within which we can consolidate "individual languages" such as Spanish or Japanese and talk about

them. Although the actual analyses we have carried out of language data remain basically the same as existing studies in linguistics, we believe it's important to consolidate the perspective and framework of the discussion which enables us in the long run to examine the possible change of the overall arrangement of what we call “individual languages”.

3. Language and Media

In order to take into perspective the possible dissolusion of “individual languages” as we talk of, it is desirable to have a brief look at the “origin” of the concept “individual language” such as Spanish, Japanese, English etc. As Saussure pointed out, “languages” (l*a langue*) within its natural environment cannot be positively consolidated as a systematic unity; diversities and variations of “languages” are continuous rather than discrete, which means geographically-oriented enclosure and consolidation of “individual languages” such as French, Italian, Spanish, German, Japanese, Chinese, etc. etc. is external to *la langue* and was in fact promoted by socio-political dynamics.

In relation to the present study, two points should be observed. First, in many cases “individual languages” such as German, French, Japanese, etc. are consolidated in parallel with the development of print media, national capital and modern nation-state. The “texts” translated from e.g. Latin or Chinese cannons are distributed in print media, which contributed to the “standardisation” (i.e. consolidation) of “individual languages”, while the geographical area of “individual languages” in many cases match nation-state due to the functioning of national capital as well as the limitation of physical communication channel through which printed texts were distributed. Secondly, this process shows that what we speak currently is inherently bound to written language (Saussure talked of the influence of “literary language” to “language”).

That “individual languages” most of us take for granted now was constructed within some specific socio-politico-media environments implies that the very arrangement of “individual languages” as well as the unity of “individual language” might dissolve if these related conditions change radically. As is pointed out by many, (a) the internet may (have) introduce(d) revolutionary changes to the arrangement of communication media, potentially as big as print media introduced, (b) the status of nation-state is in decline, although whose consequences are not necessarily desirable. In short, the quintessentially “modern” arrangement of the quatrogy of “individual language”, “print media”, “nation-state” and “national capital” is in the process of dissolving now. So it would be at least necessary to take into account the possible consequences of these to what we talk of by “individual languages” in observing the effect of new media to languages, even though —being a study of “language”— we inevitably have to start from assuming the unity of “individual language” such as Japanese, Spanish, French, English, Chinese, etc.

4. Some Points of Observation in Langauge in News Reporting in Old and New Media

In using news reporting texts (i.e. “parole” in its broader sense) to the observation of language (“la langue”), two points should be taken into account. On one hand, as it is “parole” that contribute, in the long run, to changing the nature of “la langue”, using texts is an essential condition for the task we set out in this research. On the other hand, however, actually available news reporting texts (and texts in general) are not “natural products” but the result of socio-political dynamics that is external to language. As such, rather than pretending the fake neutrality of

the available news reporting texts vis-a-vis language, it is more desirable, both theoretically and methodologically, to accept these as the basic conditions that bind the type of research we set out here. Thus in the present study, factors related to media as well as to discourse should be taken into account, even though (or precisely because) the target of the study is language itself. In the empirial analyses of textual data, therefore, at least the following points should be taken into account:

- Formal nature of different media technology: Different formal characteristics of media may require different linguistic organisations of texts. For instance, it is generally said that e-mail magazine tend to have shorter paragraphs for readability. Also, “hyperlink deixsis” is conditioned on the nature of hyperlink supported by Web etc.
- Communicational nature of different media (as social function): most of the “established” mass media, such as TV broadcast or newspapers, have “national” audience/readers as their target and a small number of “national” group as message senders; they tend to assume a putative “community”. Thus, “we” tend to refer to “uniform” people delimited by nation-state border in many cases (especially in such cases as Japan where myth of single race is widespread). On the other hand, information circulation on the Internet is quintessentially of trans-border nature, and the language can be for communicating with anybody who has sufficient command in that language, not for assumed “community”, while senders tend to be individuals or small groups that cut across established national system. This might have a far-reaching effect on linguistic expressions in news reporting, e.g. use of deixsis, vocabulary, choice of grammatical subject, theme, topic of discourse, etc.
- Closely related to these two, there is a communication nature of media technology through which the language can be affected, such as the use fo smilies, register variations, etc. (This is different from the first in that it is concerned with communication, while also different from the second in that this can affect any pairwise communication and independent of the characteristics of the media as social institutions).

From linguistic points of view (in its wider sense), the points of observation can be classified as follows (this might not be exhaustive):

- Textual characteristics: nature of textual units, length of a text, length of a paragraph, etc.
- Communicational characteristics: Use of deixsis, choice of information, subject, object, etc.
- Grammatical characteristics: Complexity of formal structure of sentences, variants of inflexions, etc.
- Lexical/lexicological characteristics: Choice of vocabulary, choice of attributes of lexical items, density of vocabulary, etc.

As a preliminary observation, we decided to focus on two aspects: (a) The use of deictic expression “Watashi” (“I”) and “Wareware” (“we”) at subject position, including their variations, and (b) the use of vocabulary by types of origin (Japanese vocabulary can be divided into four, i.e. those which are originated from Japanese, those which are originated from Chinese, those which are originated from mainly Western languages, and the mixture of these).

5. The News Reporting Data: Analyses and Observations

Five news reporting sources were used for the analysis:

NHK: “Public” TV broadcasting institution. The data is the transcription of what is stated by announcers. The data covers news stories from January 1 2003 to December 31 2003.

Mainichi (MAI): A traditional print media newspaper. The data is taken from articles between January 1 2003 to June 30 2003.

JANJAN (JAN): An online “civic” newspaper. Volunteer journalists and editors are running it online. The data is taken from October 16 2002 to May 1 2004.

Publicity (PUB): A personally-managed mail magazine. The data covers from January 2 2003 to December 30 2003.

Tanaka (TAN): A personally-managed international news cite. The data covers from January 7 2002 to April 29 2004.

We extracted reports covering “international news,” as (a) Tanaka and to some extent Publicity focus on this area, and (b) to controll domain makes the interpretations of observed phenomena more consistent (That coverage periods vary is due to the problems involved in pre-processing and not important here). Basic quantities of these five data are as follows:

	# sentences	# characters	# word tokens	# word types	# type/token
NHK	23786	1539974	911744	15668	0.0172
MAI	35471	1798307	1103081	28152	0.0255
JAN	10754	53743	334890	19167	0.0572
PUB	30112	1462039	911212	31529	0.0346
TAN	8247	544081	325791	13182	0.0405

Although type-token ratio is sample-size dependent (the smaller the sample the larger the value tends to be), it is still notable that “established” media (NHK and MAI) tend to take smaller type-token ratio. Among online and/or private media, particularly notable is PUB, whose sample size is about the same as NHK and MAI but its type-token ratio is twice the value of that of NHK.

5.1 “watashi” (first person singular) and related deixtic expressions

We observed “watashi” (first person singular) and “wareware” (first person plural) occurring in a subject (and quasi-subject) position by extracting relevant pronouns followed by subject case-marking postpositions in Japanese. We also observed “wagakuni” (my/our country). The following table shows the number of sentences that contain these expressions.

	Watashi	Wareware	Wagakuni
NHK	66 (0.28%)	120 (0.50%)	43 (0.18%)
MAI	173 (0.49%)	327 (0.92%)	51 (0.14%)
JAN	120 (1.12%)	96 (0.89%)	3 ((0.02%)
PUB	1148 (3.81%)	346 (1.15%)	22 (0.07%)
TAN	226 (2.74%)	69 (0.83%)	11 (0.13%)

Although detailed observations are needed, these figures show characteristic differences across different media. We

can draw a rough line between “established” media and new civic/personal online media. On one hand, civic/personal media tend to use first person singular (and to some extent plural) subject much more frequently than “established” media, while the former tend to use less “wagakuni” (which has a very strong connotation of “our community” whereever it is used). Observing some samples show that the majority of first person singular subject in established media are from quotations, while in JAN and PUB the reporters use the first person singular straightly. The uses of “wareware” are diverse but rough qualitative observations reveal that in established media they tend to occur in quotations while in civic/personal media they are often used as representing sender-reader community. At the moment it is too much to claim that these reflect the rearrangement of sender-community relations which we can clearly see in the messages and nature of these media themselves (personalised sender and reader community which is detached from nation-state in new media vs established sender as part of “national” community in established media), it is at least shown that deixtic expressions clearly correlate with media types.

5.2 Use of words borrowed (mainly) from Western languages

We also observed the token ratio of words borrowed (mainly) from Western languages (mostly English):

	NHK	MAI	JAN	PUB	TAN
All	40503 (4.44)	55385 (5.02)	18837 (5.62)	48879 (5.36)	22929 (7.04)
Non-PN	14220 (1.56)	24831 (2.25)	13175 (3.93)	38439 (4.22)	9901 (3.04)

Non-PN means borrowed words that are not proper names (such as person, location or organisation). This shows that, if we think of borrowed words which are not proper names (which are abundant as the data deal with international news), traditional “established” media tend to use substantially less borrowed words than online civic/personal media.

These quantitative data shows there are at least surface correlation between characteristics that manifest themselves at language level (i.e. deixis and vocabulary) and the position of media in social discourse. At the media level, it is widerly held that online personal/civic media tend to be more cross-border and do not regard the message they portray to be interpreted within the “established” nation-state framework (though naturally linguistic choice limits potential readers), while “established media” has strong relation to nation-state. In order to keep the media grounded, civic/personal new media tend to use first person singular to clarify the message sender, while in “established” media it is not regarded necessary as the “community” is assumed to be well established. On the other hand and closely correlated with this the vocabulary tend to be cross-border in online new media than traditional “established” media.

6. Conclusions

In this paper we articulated the research framework of the “new media language” with the possible dissolusion of so called “individual languages” theoretically taken into account, and reported the result of a very preliminary analyses of the actual data, using news reportings in five different media in the domain of international affairs. The results indicates the existence of the correlation between the role/position of the media and the use of language. However, we have not yet clarified the factors —on both media and language sides— that contribute to this correlation properly. For this to be done, we need to delve deeper into the historical/political/social relationships

between media and language at every level. At least, however, we have shown that this research path is promising in observing what is currently occurring in languages (not only as empirical phenomena but also as the very concept of “individual languages” as a unity).

Acknowledgement

The work reported here is supported by Hoso Bunka Foundation, Inc., Japan.

Bibliography

Aitchison, J. And Lewis, D. M. (eds) *New Media Language*. London: Routledge, 2003.

Anderson, B. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Revised edition ed. London: Verso, 1991.

Baayen, R. H. *Word Frequency Distributions*. Dordrecht: Kluwer. 2001.

Baron, N. S. “Letters by phone or speech by other means: the linguistics of email,” *Language & Communication* 18, p. 133-170, 1998.

Crystal, D. *Language and the Internet*. Cambridge: Cambridge University Press, 2001.

Febvre, L. and Martin, H-J. L*'apparition du Livre*. Paris: Edition Albin Michel, 1971.

Ferrara, K. et. al. “Interactive written discourse as an emergent register,” *Written Communication* 18(1), p. 8-34, 1991.

Karatani, K. ``Nationalism and ecriture'' *Surfaces*, vol. V.201, 1995.

Ong, W. J. *Orality and Literacy: The Technologizing of the Word*. London: Methuen. 1982.

Saussure, F. de. *Trouisieme Cours de Linguistique Generale (1910-1911) d'apres les cahiers d'Emile Constantin*. Komatsu, E. ed., Harris, R. trans. Oxford: Pergamon Press, 1993.