

Paris Lectures in Japanese Linguistics, Kurosoio Publishing Co., Tokyo
2005.

Japanese – An International Language

ANDRÉ WŁODARCZYK

Université Charles de Gaulle – Lille 3

1. Presence

As surprising as this may seem, Japanese is increasingly being used in diverse relations among different peoples. This is happening to such an extent that we can say that Japanese can already be considered a new international language. Indeed, in the last thirty years the Japanese economy has been present on the international stage and still counts – despite the financial recession of the last decade – as the second most powerful economy in the world. We do not intend to play either the role of prophet concerning the future of the Japanese language nor to take the place of the Japanese politicians whose responsibility it might be to promote their language in the world. On this point, our purpose will be limited to noting a certain number of facts.

The Japanese language began to attract attention at the beginning of the second half of the 20th century, i.e.: from the period known as “economic recovery” (1950-59) after the Second World War. As evidence of this the world production of books in 1952 stood at 22% for English, 16% for URSS languages, 15% for German, 12% for Japanese, 10% for French, 7% for Spanish etc. Even although the Japanese translate from many foreign languages, their language is not among those that were the most translated during that period (cf. Burney P., 1966, p. 68-69). Today, it is noteworthy that Japanese is increasingly becoming a true

international language. There are many signs of this tendency. For example, in Europe, translations are most often carried out from English but Japanese is – this is not a very well known fact – the second language from which the member countries of the European Union translate the most often. Here, again, are some figures: (a) the most translated source languages: English – 48%, Japanese – 23%, French – 8%, German – 5%, Russian – 2%, others – 5% and (b) the target languages towards which translations are carried out the most: English – 45%, Japanese – 24%, French – 12%, Spanish – 10%, German – 5%, others – 4%. A further example: a great increase in the number of students (and proportionally of teachers) of Japanese in the world (out with Japan) has been noted: in 1974 – less than 100 thousand and in 1990 – nearly one million.

At the beginning of the nineties, the Japanese ministry of education (Mombusho) took an interest in the problem of internationalising Japanese. Indeed, having noted the large increase in the number of people learning Japanese abroad (between 1979 and 1993, this number was multiplied by 13), the Mombusho decided to hold an inquiry into the measures to be taken to promote Japanese teaching.

In 1993, the Mombusho published a report with an eloquent subtitle: “Towards the internationalisation of Japanese”. Two large-scale projects were created in 1994: one placed under the aegis of the Japan Foundation, which until then was solely responsible for the promotion of Japanese in the world and another depending directly from the Mombusho. Moreover, the JF organised, in collaboration with three European organisations (the Franco Japanese Institute of Tokyo, the Goethe-Institute of Tokyo and the British Council) a conference called “The new needs of international communication and language teaching”. During this conference (Tokyo, 16-17 Oct. 1993), several measures were announced concerning the increase both of teaching centres and of work force. In France, Japanese is currently taught to more than 7,500 people (secondary and higher education together, [Origas J.-J. 1994]). In the USA, several universities devote “pages” on their Internet sites to the teaching of Japanese. At MIT, for example, with Canon and the Consortium for Language Teaching and Learning’s assistance, a computer service called JP NET has been inaugurated. This service provides both teachers and students with documentation born of research carried out earlier by Canon as part of its Athena project (automatic translation system), and particularly the information concerning vocabulary, structural exercises, grammar, ideas for communication activities etc.

Japanese’s move to the forefront of the international stage corresponds to research into the automatic processing of the Japanese language (particularly in automatic lexicology and computer-aided translation). This is due to the fact that the dematerialised information (recorded in the structured memory of the computer systems) requires increasingly sophisticated processing of languages rather than simply

being recorded in writing or on tape. Only three Japanese government initiatives will be mentioned here (whilst keeping in mind that the co-ordination network in the scientific and technical research field lies with the private industrial sector (76%) and that the Japanese State subsidises 18% of the research, the remaining 6% belonging to universities). They are initiatives created by the State, manifestly, as priority projects in Japanese politics in the last two decades:

- an automatic translation project called “Project Mu” (between 1983 and 1992) which was part of a series of projects aiming to create computers able to understand human language. The goals of the project were to create (1) a Database (dictionary of scientific terminology), (2) Software: a device to convert syntactic structures, (3) a Translation System combining both. Only English and Japanese were concerned,
- two automatic lexicology projects (the EDR – Institute for electronic dictionary research was created in 1986 along with the ICOT (Institute for new generation computers) and the IPA (Agency for promoting computer technology). The first of these projects came to an end in 1994 (with satisfactory results – 13 CD-ROMs of data) while the second has been in progress since 1987.

At present, most of these projects are re-orienting towards research, processing and stocking information. Finally, it is important to note that new linguistic theories of Japanese appeared, in the wake of and profiting from the experiments in the automatic processing of Japanese. The two most remarkable models are GUNJI Takao and MIZUTANI Shizuo’s grammars. We will come back to these problems in the Research paragraph.

2. Genesis and Structure

Japanese is also becoming an international language despite certain linguists’ affirmations that this will never happen for structural reasons. Indeed, Japanese has a very different structure to western languages, all the more so as Chinese, a language which belongs to yet another linguistic family, has had a great influence not only on the written system of Japanese, but also on its grammar and vocabulary. It is for this reason, no doubt, that Japanese has become a complex language whose learning is very difficult for foreigners. Moreover, Japan has its own long philosophical tradition concerning writing and grammar, which sometimes makes it difficult to understand the concepts which linguists and computer scientists are manipulating today.

Japanese has poorly established relations with other languages (but cf. Catherine Garnier, in the present volume) although comparative methods in historical linguistics hypothesise that relations with Korean should be possible, all the more so as neither of the two languages have close relations which are known elsewhere. Glotto-chronological calculations carried out in Japan are not convincing either, as the results obtained, which suggest that the separation of Japanese and Korean took

place around 5,500-5,800 BC, are very close to the limits imposed by the powerfulness of the formulae used (bordering chance). Nevertheless, note that the glotto-chronological method provides converging results concerning the calculation of the age of different Japanese dialects (Hattori S. 1959).

We do not know when Japanese began to be spoken on the archipelago of Japan. However the presence of Japanese anthroponyms and toponyms in Chinese chronicles from the 4th century, as well as the appearance of Japanese inscriptions on stone and metals, in ideographic characters borrowed from China in the middle of the 5th century, seem to confirm the results of the calculations since they show that the two types of dialects of the Ryūkū and Hondo islands were separated from their mother language at the same period.

Phonology. The phonological composition is very simple. Standard Japanese has the following 22 phonemes: 5 vocalic phonemes (a,e,o,i,u), 13 consonantal phonemes: (p,b,m,t,d,n,s,z,k,g,ng,r and h), 2 semi-vowels (bilabial: w and palatal: j) and two special phonemes (written in an abstract way as N and Q). The last two phonemes have a mora quality (phonic unit with consonantal structure but only two times shorter than the phonic unit with vocalic structure), which means the syllable - whose structure is also simple, either open (CV) or containing moraic geminates between two syllables (CV[N|Q]CV) – is not the only rhythmic unit in Japanese. Therefore, the Japanese syllable does not have consonant groups at the beginning (CCV is impossible) and is always open (VC is also impossible except for those which end with mora-phonemes). The Japanese accent is a height accent with a high and low register. The accent hits a mora or a series of moras in the framework of an accented word.

Morphology. Japanese morphology is of the mixed type: flexion-agglutinated. Just as the lexical units subdivide into two classes, those of *taigen* (invariable words) and of *yōgen* (variable words), brought to light by dichotomy of variability criteria, the grammatical units subdivide also, into two morphological categories called *joshi* (auxiliary terms) and *jodōshi* (conjugated auxiliary terms). Given that the case morphemes are invariable and bound, Japanese grammatical tradition has not developed a concept of declension. Moreover, Japanese conjugation differs greatly from Indo-European languages' conjugation. Indeed, the paradigms of Japanese conjugation are doubly composed possessing both a median part (called "auxiliary") and a final part (called "formal"). Among the verbal categories, one can find time (anterior / non anterior), aspect (accomplished / non accomplished), diathesis (active / passive / causative), politeness (deferent / honorific / deprecative), negation, desiderative etc. The person category is not verbal in Japanese, although a system of personal pronouns does exist (a very open one, incidentally). This can be explained by the fact that the Japanese-speaking subject does not express his identity directly but in relation to the other utterance actors.

Syntax. Subject omission is possible in Japanese because this language does not have the morphological agreement which characterises the Subject-Predicate relation in all the Indo-European languages. The optional presence of the utterance's subject sometimes allows Japanese grammarians to conclude an absence of Subject-Predicate structure in the language. This made the task of defining predication difficult. The latter, often confused with the utterance, has even given place to a theory known in Japan as chinjutsu-ron (theory of predication / utterance). Japanese has no relative pronouns, as the subordinate term always precedes the principal term. In the same way in determination, the determining term precedes the determined term. One of the particularities of Japanese syntax is the great number of grammatical units which refer directly to the utterance. This fact is often cited as proof of the specificity of Japanese utterance.

Lexicon. The Japanese lexical inventory is doubly structured (sino-japanese). Moreover, the morpho-syllabic units of Chinese origin, functioning in a system whose elements are made up of indigenous polysyllabic units, present a high degree of lexical productivity. Since the introduction of science and western techniques to Japan, Japanese vocabulary has become considerably richer with false borrowings from Chinese, "false" - because they are words created by the Japanese from units borrowed from Chinese. Consequently, today such "borrowed" vocabulary used by the daily press reaches 52% in comparison to the 45% of words of Japanese origin, 1% of words of other origins than Chinese and 2% of words made up of heterogeneous elements.

Writing. Japanese uses three writing systems, which coexist syncretically. Notably a pseudo-ideographic system of Chinese origin (kanji) and two syllabaries (kana) which contribute to all notations of Japanese. Both syllabaires are derived from Chinese characters used in the role of phonograms in Japanese: hira-gana - to note the grammatical units and some indigenous words and kata-kana - for noting foreign borrowings (Chinese borrowings aside). The first latin transcriptions of Japanese date from the end of the 16th century when Portuguese missionaries started to study the language. Subsequently, first the Dutch, then the Germans, the French and the Americans developed their respective transcriptions of Japanese based on latin script.

Encoding and processing kanji. Due to the extreme difficulty of classifying kanji in a rational manner, the different computing systems in Japan today use three different codes. Given their great number, the kanji characters are coded on two octets. For example, in the hexadecimal code Shift-JIS, the first octet is considered to be the high octet, the second the low octet. The Shift-JIS code is therefore the result of a combination of ASCII code units. This is the reason why, when text processing is not compatible with kanji, users see mojibake (unformed characters) on their screens.

3. Linguistic research

Japan has a rich tradition of reflecting on language (cf. Włodarczyk A. - 1982). For this reason, "linguistics" today has several equivalents: kokugogaku (studies of the national language), (gaikoku-)gogaku (studies of (foreign) languages), gengogaku (studies of language, from which ippan-gengogaku - general linguistics). Although all the currents, which animated linguistic studies in the West, reached the Nippon archipelago nearly a century ago, traditional linguistics survives, more than elsewhere, beside the modern linguistic theories. Let us mention however two modern linguistic schools of which one (1°) is inspired by western linguistics but tries to adapt its concepts and methods to the realities of the Japanese language thus giving birth to inevitably original solutions - Gunji T. (1987) and the other (2°) only borrows the logic-mechanic tools to systemise traditional linguistic knowledge - Mizutani S. (1991). GUNJI Takao profited from the results of research devoted to English using the unifying grammars (notably HPSG) to describe a set of facts of the Japanese language in a very innovative way. His grammar, the original of which was published in English, is known by its acronym JPSG. Like its big sister, it enables the processing of languages whose linear order is relatively free. It is interesting to note that the ICOT, in the last stage of its existence, used JPSG during the development of its version of Prolog (cu-Prolog) - the first programming language with symbolic constraints using Horn clauses. On the other hand, MIZUTANI Shizuo built a formal model for Japanese grammar based on concepts elaborated by the indigenous linguistic tradition. It concerns of course, and above all, syntactic phenomenon, but the particularity of this grammar is that the components are very original and the structures are not only trees but also, sometimes, trellises. We thought fit to make this grammar accessible to Western linguists by translating it into French (Mizutani S., 1991), but already a new presentation of the grammar, founded on the theory of sets, has seen the day in Japan.

Let us mention, finally, that these new linguistic theories of Japanese appeared in the wake of and profited from the experiments on the automatic processing of the language (cf. Ishiwata T. & Włodarczyk A., 1997). Indeed, Japanese - like the other international languages - has to be computerised. This is not easy since, as we said previously, it is not related to any of them. The ultimate goal (which is by far the only practical aim) of the natural language processing is machine translation, and it is precisely in this domain that research is the least advanced. No doubt, this is because the automatic translation of Japanese into a European language requires the computer to memorise an enormous quantity of knowledge (as well as grammar rules). The most spectacular results have been obtained in the ATR research institutes in Kyoto for the needs of telephony by doing demonstrations of conversation translations. Presented in August 1992 in Nantes at the 15th international conference of computational linguistics COLING-92 (20-28 July 1992) the

Japanese system recognises Japanese utterances entered by microphone and generates their English equivalents by loudspeaker while posting some intermediary results of analysis and synthesis in a very convivial environment (several multicoloured windows). But it is clear that computer-processing of Japanese encounters many obstacles of the structural type, which are of a different nature to those encountered in the research on western languages. The large structural difference, which characterises Japanese in comparison to European languages, is also the cause of many difficulties in automatic translation. Let us take for example the fact that in Japanese there is no distinction between singular and plural. This makes the translation from Japanese to European languages difficult. Then again, to take another example, the fact that Japanese differs greatly from other European languages in its use of pragmatic maxims such as politeness. Although this problem is being largely studied in an experimental way at the moment (for example IBM-Japan's recent project SHALT2), none of the commercialised systems take it into consideration; generally stopping at syntactic and pseudo-semantic processing.

Nevertheless, in fundamental computer science in Japan, the questions concerning the automatic processing of Japanese remain topical in the creation of all research projects. Indeed, fifteen years ago, the Japanese government spectacularly decided to promote computer research (notably, by creating in 1982 the Institute for the new generation of computers - ICOT). According to the conception of what was called "the fifth generation of computers", the computer system must be equipped with artificial intelligence and knowledge including linguistic capacities. In other words, the idea of future computers speaking natural language seems to be a given in Japan's scientific and political classes. Witness the research carried out at the ICOT itself and for 12 years in the ICOT annex: Institute for the research into electronic dictionaries whose 13 CD-ROM are used today by several institutes in their own research and development projects. Witness also the research carried out in the ETL laboratories (Laboratories of electronic technologies, RWC "Real World Computing" project) and ATR (Institute of research of advanced telephony, automatic translation of the spoken language with speech recognition project).

4. Promotion and teaching.

As soon as the international need is felt and the prestige of the country's national language increases, the necessity to put into place a linguistic policy emerges, which consists in the first place of promoting the teaching of the language in question not only as a mother tongue but also as a foreign language.

"One must first of all admit that the primary reasons for a linguistic policy, which applies itself to favouring the spreading of its language abroad, are principally political and economic." (Blaasch Hans-Werner, 1993).

In 1985, CDI company – Sôgô Kenkyû Kaihatsu Kikô (Institute for Development and Research – CDI) published its report of enquiry (814 pages – NIRA OUTPUT) concerning the teaching and the propagation of Japanese. The starting point of this enquiry was the realisation that increasing numbers of foreigners were learning Japanese both in Japan and abroad. The enquiry revealed amongst other things the varied motivations given by the foreigners. Seven in all: (1) scientific and technical progress, (2) economic development, (3) tourism, (4) international exchanges, (5) cultural tradition, (6) family relations and (7) particular interests and leisure. Europe is mentioned especially for its cultural tastes but people are interested in Japan particularly because it is progressing on the modernisation path. The same year, the Japan Foundation (Kokusai Kôryû Kikin), which had been created in 1972 to “promote the cultural exchange between Japan and other countries”, published its 7th volume in the series of directories entitled this time “Japanese Studies in Europe”.

In 1990, the same Japan Foundation, which incidentally – in its Japanese name – includes the word “international” (kokusai), announced [JF Newsletter, August 1990] its initiative to publish a review under the significant title “teaching of Japanese in the world” (Sekai no nihongu kyôiku). Today, there are also several regular publications for Japanese as a foreign language-teaching in Japan. In 1994, two large-scale projects were born: one placed under the aegis of the same Japan Foundation (which until then in fact was solely responsible for promoting Japanese in the world) and another answering directly to the Mom-busho. Moreover, the Japan Foundation organised, in October 1993, in collaboration with three European organisms (the Franco-Japanese Institute of Tokyo, the Goethe-Institute of Tokyo and the British Council), a conference entitled “the new needs of international communication and language teaching”. During this conference, several measures were announced concerning both the increase of teaching centres and workforce. It is interesting to note that in 1990 France was in 10th position with 7,746 learners and 271 teachers.

The AREA review of 3rd April 1995 published the following figures: during the period 1979-93, the number of foreigners learning Japanese was multiplied by 13 (this number increased from 127,000 in 1979 to 1,623,000 in 1993). It goes without saying that this inspired a series of actions for teacher-training as well as the writing of new manuals and the creation of other materials and pedagogic tools, contributing in this way to the diversification of methods: not only must Japanese culture, in its strict sense, be taught, but also everything which concerns life (the individual's, the family's, the society's), thought, communication, economy and politics.

At the moment, around the world, modern methods are used (satellite television, multimedia on CD-ROM and by internet etc.) in the teaching of Japanese. Above all an example of this tendency is the construction of the JP NET network described as “a virtual global com-

munity of specialists of Japanese studies” and launched in 1995 by the Massachusetts Institute of Technology.

Nowadays, no language is fully international on a planetary scale; even English is not understood in the same way everywhere. On the other hand, there are geographic areas where such and such a language is used more often (or more voluntarily) than such and such other language. It seems that Japanese in all likelihood could serve, more so than today, as an intermediary in the relations between the different Pacific area countries.

5. References

- BLAASCH Hans-Werner, (1993) “La diffusion des langues nationales en tant que langues étrangères à l'étrangers”, in conference proceedings “Nouveaux besoins en communication internationale et enseignement des langues”, (Tokyo, 16-17 oct. 1993)
- BURNEY Pierre (1966) Les Langues internationales, "Que sais-je" Collection, PUF, Paris
- Collective document (1990) Announcement concerning the creation of the review «*Sekai no nihongo kyōiku*» (Teaching Japanese in the world), The Japan Foundation, Newsletter, Tokyo
- GUNJI Takao (1987) Japanese Phrase Structure Grammar - An Unification-Based Approach, Series Studies in Natural Language & Linguistic Theory, D. Reidel Publishing Company, Dordrecht-Boston-Lancaster-Tokyo
- HATTORI Shirô (1959) *Nihon-go no keitô* (Parenté du japonais), Iwanami shoten, (1st edition), Tokyo
- ISHIWATA Toshio & WLODARCZYK André (1997) “Le Japonais et l'ordinateur - présentation générale”, in «Travaux de linguistique japonaise» (University of Paris 7 publication), volume 10 - “Langue - Ordinateur - Mentalité”, Paris
- MIZUTANI Shizuo (1991) “Description Systématique de la Grammaire Japonaise”, trans. by Reiko Shimamori and André Włodarczyk (pp. 15-175), followed by “Application de la grammaire de Mizutani au traitement informatique” by André Włodarczyk, (pp. 177-198), in «Travaux de linguistique japonaise», Vol. IX, University of Paris VII, Paris
- ORIGAS Jean-Jacques (1994) “Furansu ni okeru Nihongo-kyōiku to kadai” (The state and problems of teaching Japanese in France) in «Current Report on Japanese-Language Education around the Globe», vol. 1, The Japan Foundation (pp. 173-176), Tokyo
- Collective work (1985) NIRA OUTPUT NRC-83-2 *Nihongo kyōiku oyobi Nihongo fukyû-katsudô no genjô to kadai* (The current state and the problems of teaching Japanese and the activities aimed at

promoting the Japanese language), CDI - Sôgô Kenkyû Kaihatsu Kikô (Institute for Development and Research - CDI), pp. 814, Tokyo

Collective work (1993) Conference proceedings “Nouveaux besoins en communication internationale et enseignement des langues”, (Tokyo, 16-17 oct. 1993), The Japan Foundation, Tokyo

TSUDA Hiroshi (1994) “Cu-Prolog for Constraint-based Natural Language Processing”, in IEICE Transactions on Information and Systems, vol. E77-D, N° 2, Tokyo

WLODARCZYK André (1982) “Théories du langage au Japon”, in «Langages» (trimestriel review), No 12/68, LAROUSSE Editions, Paris

PARIS LECTURES IN JAPANESE LINGUISTICS

Japanese – An International Language (André WŁODARCZYK) 1

Part 1. Japanese Language in General Linguistics

Japanese as Seen from Outside and from Inside (Yoshihiko IKE-GAMI)	13
From Japanese to General Linguistics - starting with the wa and ga particles (André WŁODARCZYK)	25
From Rodriguez to Computational Linguistics – Japanese in European Linguistics (Viktoria ESCHBACH-SZABO)	60
Will Cognitive Change Cause GA to O Change In Constructions Of GA/O Alternation? (Seiichi MAKINO)	76
Between Grammaticalization and Drift (Lone TAKEUCHI)	94

Part 2. Language Teaching Methodology

Integration of Language and Culture (Hideo HOSOKAWA)	131
--	-----

Part 3. Computational approaches

A Multi-modal Approach to Lexical Semantics (Naoyuki OKADA)	143
An Entity-Based Linguistic Framework: the Case of Japanese Syntax (Zoya M.SHALYAPINA)	169
Information Reception and Dispatch using Internet Machine Translation Systems (Shoichi YOKOYAMA)	212