



Using corpora to find relevant examples of Japanese honorific forms in order to generate exercises for intermediate learners of Japanese



UNIVERSITÉ DE STRASBOURG

**Corpus et Outils
en linguistique, langues et parole:
Statuts, Usages et Mésusages**

Using corpora to find relevant examples of Japanese honorific forms in order to generate exercises for intermediate learners of Japanese

Valérie Collec-Clerc (www.valtal.fr)

(mail: valerie.clerccollec@yahoo.fr)

LABORATOIRE
D'INFORMATIQUE
FONDAMENTALE
de Marseille
www.lif.univ-mrs.fr





**Using corpora to find relevant examples of Japanese honorific forms
in order to generate exercises for intermediate learners of Japanese**

- Research context
- Why do we need corpus analysis ?
- Corpora and tools
- Adopted methodology for the use of corpora
- Future research perspectives



**Using corpora to find relevant examples of Japanese honorific forms
in order to generate exercises for intermediate learners of Japanese**

- **Research context**
- Why do we need corpus analysis ?
- Corpora and tools
- Adopted methodology for the use of corpora
- Future research perspectives



**Using corpora to find relevant examples of Japanese honorific forms
in order to generate exercises for intermediate learners of Japanese**

■ **Research context**

- Goal: Generating sentences and exercises for intermediate learners of Japanese
- Characteristics of Japanese
- Japanese politeness language in a bird's eye view



Using corpora to find relevant examples of Japanese honorific forms in order to generate exercises for intermediate learners of Japanese

- **Research context**

- **Goal: Generating sentences and exercises for intermediate learners of Japanese**
- Characteristics of Japanese
- Japanese politeness language in a bird's eye view



Using corpora to find relevant examples of Japanese honorific forms in order to generate exercises for intermediate learners of Japanese

Research context

Goal: Generating sentences and exercises for intermediate learners of Japanese

- Learning aids for the Japanese learners through exercises on sentence structures (like “phrasebook” approach already used for beginners) : **Zock M.** and **Lapalme G. 2010 - *A Generic Tool for Creating and Using Multilingual Phrasebooks.***
 - Target learners : intermediate level ↔ syntactically complex sentences which rely on semantic and contextual elements (e.g.: Japanese formal politeness language, also called the “honorific” system).
- Identifying and formalising “rules” or standard patterns to generate realistic sentences **It does not only consist of syntactic rules but also relies on the semantic/pragmatic constraints of the situation of uttering.**



**Using corpora to find relevant examples of Japanese honorific forms
in order to generate exercises for intermediate learners of Japanese**

- **Research context**

- Goal: Generating sentences and exercises for intermediate learners of Japanese
- **Characteristics of Japanese**
- Japanese politeness language in a bird's eye view



Using corpora to find relevant examples of Japanese honorific forms in order to generate exercises for intermediate learners of Japanese

Research context

Characteristics of Japanese

- Writing systems :
 - Kanji : 自然言語処理 (nouns, verb stems)
 - Syllabic scripts
 - Hiragana しぜんげんごしより (furigana, grammatical words, inflections)
 - Katakana (フランス) (foreign words)
 - Latin script (romaji) (shizengengoshori)
 - No space between words
 - Word order (SOV) with case particles after the nouns
 - No grammatical gender and number for nouns
 - Examples : フランス人の学生が自然言語処理を勉強する。
- Linguistic tools mainly designed for Western languages are not adapted to Japanese**



**Using corpora to find relevant examples of Japanese honorific forms
in order to generate exercises for intermediate learners of Japanese**

- **Research context**

- Goal: Generation of sentences and exercises for intermediate learners of Japanese
- Characteristics of Japanese
- **Japanese politeness language in a bird's eye view**



Using corpora to find relevant examples of Japanese honorific forms in order to generate exercises for intermediate learners of Japanese

Research context

Japanese politeness language in bird's eye view 1/5

- Focus = formal politeness system (generally regarded as complex for non native speakers)
- The sentence construction used reflects the location of speaker inside the Japanese inter-personal communication system (Taiguhyougen 待遇表現)
 - Vertical relationship (jougekankei : 上下関係)
 - Individual : superior/inferior (目上 (meue)/目下 (meshita))
 - Group (more important/ important)
 - Horizontal relationship
 - In-group (uchi : 内) (honne : 本音)
 - Out-group (soto : 外) (tatemae: 建前)
 - Psychological distance (shinso : 親疎)

→ Contextual facts cannot be ignored to build up a correct sentence



Using corpora to find relevant examples of Japanese honorific forms in order to generate exercises for intermediate learners of Japanese

Research context

Japanese politeness language in a bird's eye view 2/5

Addressee-oriented politeness

- Plain form (futsukei : 普通形)
 - Neutral written language (日本語を学ぶ、きれいだ)
 - Familiar spoken language
- Polite form or Addressee honorifics (Teineikei : 丁寧形)
 - Showing respect towards the addressee
 - (masu: ます => 日本語を学びます) (desu: です => きれいです)
 - Exalted language : Emphasizing the speakers' willingness to show respect towards the addressee
 - (gozaimasu : ございます => きれいでございます);
 - (teorimasu : ております)



Using corpora to find relevant examples of Japanese honorific forms in order to generate exercises for intermediate learners of Japanese

Research context

Japanese politeness language in bird's eye view 3/5

Referential politeness

Often referred as *keigo* : 敬語

- **Subject honorifics, called *Sonkeigo***: (尊敬語)

The speakers elevate or show respect towards the subject of the utterance (appreciative towards the non speakers).

- **Non-subject honorifics, called Humility, or *Kenjougo***: (謙讓語)

The speakers humble themselves by showing respect to the non-subject referent, generally the object of the utterance (depreciative towards the speakers).



Using corpora to find relevant examples of Japanese honorific forms
in order to generate exercises for intermediate learners of Japanese

Japanese politeness language in bird's eye view 4/5

Subject honorifics – Main verbal construction

- Passive form
 - (RARERU/RERU) : 今朝の新聞をもう読まれましたか。
 - kesanoshimbun wo mou yomaremashitaka ?
- Honorific construction
 - O +[STEM FORM]+ NI NARU
 - 今朝の新聞をもうお読みになりましたか。
 - Kesanoshimbun wo mou oyomi ni narimashita ka?
- Specific honorific verbs
 - For instance : the use of the verb いらっしゃる (IRASSHARU) (come, go) instead of 来る *kuru* (come) or 行く *iku* (go)



Using corpora to find relevant examples of Japanese honorific forms in order to generate exercises for intermediate learners of Japanese

Research context

Japanese politeness language in bird's eye view 5/5

Non-subject honorifics : Main constructions

- ageru (上げる): as a completive auxiliary
 - 申し上げる (言う: say) 差し上げる (上げる: give)
- o-verbal stem + SURU/ITASU (お読みする oyomi suru : read) (お返し致す okaeshi itasu : return)
- go-verbal noun + SURU (ご案内する goannai suru : show around, guide)
- 拝 Hai (worship) + sino-japanese verb reading + SURU
 - 拝見 (haiken) する (見る miru : see)、拝借 (hai shaku) する (借りる kariru : borrow)



**Using corpora to find relevant examples of Japanese honorific forms
in order to generate exercises for intermediate learners of Japanese**

- Research context
- **Why do we need corpus analysis ?**
- Corpora and tools
- Adopted methodology for the use of corpora
- Future research perspectives



Using corpora to find relevant examples of Japanese honorific forms in order to generate exercises for intermediate learners of Japanese

Why do we need corpus analysis ?

- Existing formalising approaches = dedicated to a specific approach
 - The speaker
 - The hearer (the addressee)
 - The context of utterance

- Deeper study of the honorific forms taking in account the most relevant documents

- Identification of the most frequently used honorific forms within the same document gender



**Using corpora to find relevant examples of Japanese honorific forms
in order to generate exercises for intermediate learners of Japanese**

- Research context
- Why do we need corpus analysis ?
- **Corpora and tools**
- Adopted methodology for the use of corpora
- Future research perspectives



**Using corpora to find relevant examples of Japanese honorific forms
in order to generate exercises for intermediate learners of Japanese**

■ **Corpora and tools**

- Segmentation problem
- Selected Corpora and tools
- Open search on the web

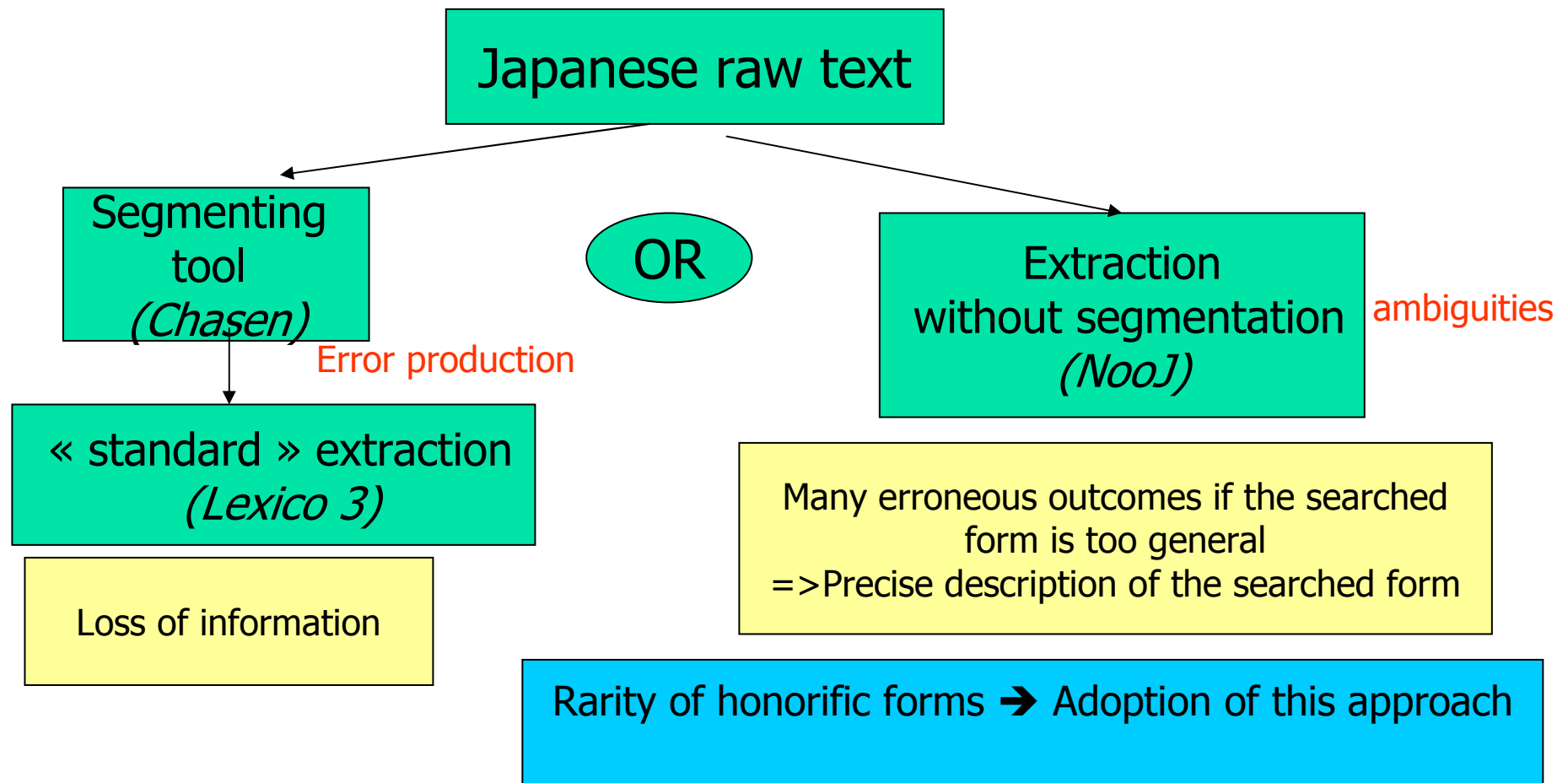


**Using corpora to find relevant examples of Japanese honorific forms
in order to generate exercises for intermediate learners of Japanese**

- **Corpora and tools**
 - **Segmentation problem**
 - Selected Corpora and tools
 - Open search on the web

Using corpora to find relevant examples of Japanese honorific forms in order to generate exercises for intermediate learners of Japanese

Corpora and tools: segmentation problem



- **Corpora and tools**
 - Segmentation problem
 - **Selected corpora and tools**
 - **BCCWJ (corpus)**
 - **SAGACE (corpus + tool)**
 - **NooJ (tool)**
 - Open search on the web



Using corpora to find relevant examples of Japanese honorific forms in order to generate exercises for intermediate learners of Japanese

Corpora and tools

BCCWJ: Balanced Corpus of Contemporary Written Japanese

- BCCWJ is a corpus computed between 2005 and 2011, as a contemporary part of the KOTONOHA corpus of written Japanese. It contains more than 100 million words.
- Contents has various origins: but the main part was generated by a systematic sampling of books, newspapers, and official texts,
- It is organized in sub-corpora (texts 2001-2005, books from 1985, best-selling books, web, governmental papers, laws, ...)
- On line access is available, to search KWiC. XML CDROM can also be obtained.



**Using corpora to find relevant examples of Japanese honorific forms
in order to generate exercises for intermediate learners of Japanese**

Corpora and tools
SAGACE

- SAGACE : a tool to extract morphological forms from non-segmented texts for Asian languages.
- On line tool is directly usable with Japanese corpora from several sources and genres.



Using corpora to find relevant examples of Japanese honorific forms in order to generate exercises for intermediate learners of Japanese

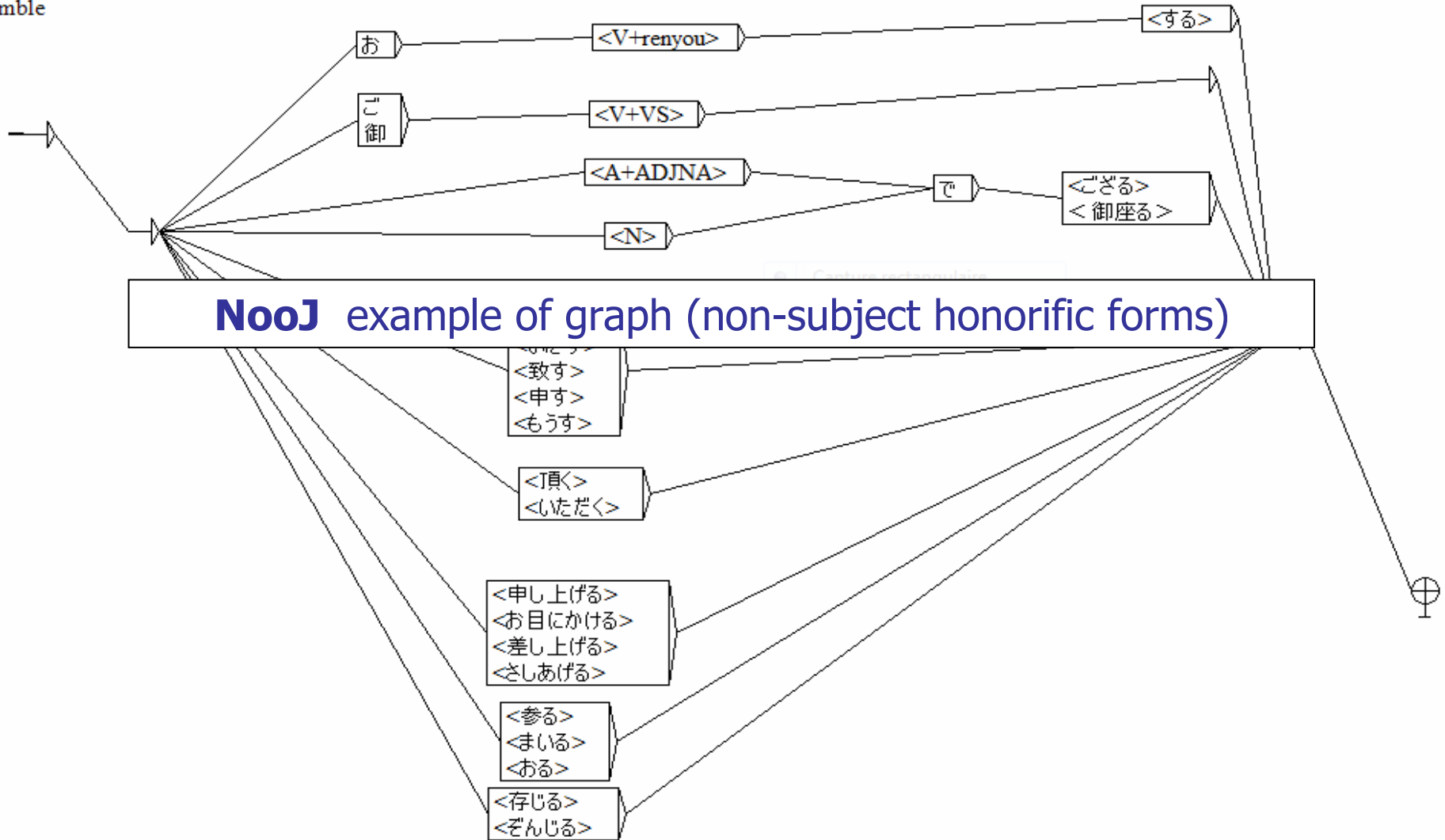
Corpora and tools NooJ

- Nooj enables word extractions from non-segmented texts and displays the results in a KWIC (Key Word in Context) environment.
- NooJ makes it possible the creation of graphs which describe parts of speech, inflectional and derivational forms (with the help of lexicons).
- Disadvantages : Necessity to implement these resources if not already existing
 - **Creation of linguistic resources for NooJ (Lexicon, inflectional and derivational rules)**



Using corpora to find relevant examples of Japanese honorific forms in order to generate exercises for intermediate learners of Japanese

humble



Using corpora to find relevant examples of Japanese honorific forms in order to generate exercises for intermediate learners of Japanese

開 ○国土計画局総務課長 いただきます。

会

それでは、ただいまから国土審議会第8回調査改革部会を開催させていただきます。

私は、国土計画局総務課長の石井でございます。本日の司会を務めさせていただきます。会議の冒頭にあたりまして、本会議の公開についてご説明申し上げます。国土審議会運営規則により、会議は原則として公開することとなっておりますので、前回と同様、本日の会議は、一般の方々にも傍聴をいただいております。今日お三方が遅れていらっしゃいますが、お三方の方も含め、以降の議事進行につきましては、中村部会長から申し上げます。○中村部会長 それでは、本日の議事に入りたいと思います。本日の議事次第はお手元にあるとおりでございますが、1つは、この審議会での懸案事項でありました「国土形成計画法について」でございます。2つ目が「今後の国土政策の方向と主要な課題に係る論点について」議論をいたしたいと思います。なお、今後の国土政策の方向と主要な課題に係る論点については、前回に引き続き、事務局資料をもとに幅広くご議論をいただければと思います。それでは、事務局よりご報告をお願いいたします。○大臣官房参事官 す。議事の1の「国土形成計画法」につきまして、私から関係の資料をご説明申し上げます。国土形成計画法の関係の資料につきましては、資料1の枝番で1～4まででございます。若干順番は前後いたしますけれども、まず、資料1-3をお開きいただければと存じます。既に3月の本部会にもご報告を申し上げますけれども、従前の国土総合開発法を改正する形で国土形成計画法を今国会に成立を見たわけでございます。ごくごく簡単におさらいをさせていただきます。真ん中の辺りでございますが、従前の全国総合開発計画を国土形成計画と改めまして、全国計画と広域地方計画、この2層で計画を進めているということでございます。その際、そのすぐ下の水色のところがございますように「計画への多様な主体の参画」ということで、国への計画提案制度、国民の意見を反映させる仕組み、諸々の新しい工夫を講じております。また、計画の理念という観点につきましても、下半分でございますが、従前のとすれば、「開発」基調、量的拡大という思想から、新しい法律には、計画の理念を、成熟社会型の計画ということで、ここにありまするような、景観、環境、有限な資源の利用・保全、あるいは既存ストック、あるいは国民生活の安全・安心といったことを書き込みまして、新たな計画体系への転換を図ったところでございます。資料1-1にお戻りをいただきたいと思います。国土形成計画法の国会の審議の経過につきまして1参事官をいたしております栗田と申します。どうぞよろしくお願い申し上げます。

Using corpora to find relevant examples of Japanese honorific forms in order to generate exercises for intermediate learners of Japanese

Concordance for Text kaikaku_gijiroku.not

Reset Display: 50 characters before, and 50 after. Display: Matches Outputs

Before	Seq.	After
。本日の司会を務めさせていただきます。会議の冒頭にあたりまして、本会議の公開についてご説明	申し上げます/Depr	。国土審議会運営規則により、
般の方々にも傍聴をいただいております。この点につきまして、あらかじめご了承くださいませようお願	申し上げます/Depr	。なお、今日お三方が遅れていら
い点につきまして、あらかじめご了承くださいませようお願申し上げます。なお、今日お三方が遅れて	いらっしゃいます/Appr	が、お三方の方も含めまして定足
お三方が遅れていらっしゃいますが、お三方の方も含めまして定足数を満たしておりますので、そのことを	申し/Depr	添えさせていただきたいと思
事進行につきましては、中村部会長にお願いしたいと思います。それでは、中村先生、よろしくお願	いたします/Depr	。○中村部会長 それでは、本日
いについて」でございます。2つ目が「今後の国土政策の方向と主要な 課題に係る論点について」議論を	いたしました/Depr	いと思
事務局資料をもとに幅広くご議論をいただければと思		いと思
い。○大臣官房参事官 す。議事の1の「国土形成計画法		す。国土形成計画法
形成計画法の関係の資料につきましては、資料1の枝番で1～4まででございます。若干順 番は前後	いたします/Depr	けれども、まず、資料1-3をお開
1～4まででございます。若干順 番は前後いたしますけれども、まず、資料1-3をお開きいただければと	存じます/Depr	。既に3月の本部会にもご報告を
いたしますけれども、まず、資料1-3をお開きいただければと存じます。既に3月の本部会にもご報告を	申し上げて/Depr	おりますけれども、従前の国土総
りますけれども、従前の国土総合開発法を改正する形で国土形成計画法を今国会に成立を見た	わけでございます/Depr	。ごごく簡単におさらいをさせ
る計画を国土形成計画と改めまして、全国計画と広域地方計画、この2層で計画を進めているという	ことでございます/Depr	。その際、そのすぐ下の水色のとこ
決させる仕組み、諸々の新しい工夫を講じております。また、計画の理 念という観点につきましても、下	半分でございます/Depr	が、従前のともすれば、「開発」基
るいは国民生活の安全・安心といったことを書き込みまして、新たな計画体系への転換を図ったという	ところでございます/Depr	。資料1-1にお戻りをいただき
って、新たな計画体系への転換を図ったというところでございます。資料1-1にお戻りをいただきたいと	存じます/Depr	。国土形成計画法の国会の審議
1-1にお戻りをいただきたいと存じます。国土形成計画法の国会の審議の経過につきま 1 参事官を	いたして/Depr	おります栗田と申します。どうぞよろ
きたいと存じます。国土形成計画法の国会の審議の経過につきま 1 参事官をいたしております栗田と	申します/Depr	。どうぞよろしくお願
い計画法の国会の審議の経過につきま 1 参事官をいたしております栗田と申します。どうぞよろしくお願	いたし/Depr	ま して、ご報告をさせていただ
い本部会の審議委員に参考とということも、国会の場におきまして、ご意見の陳述を行っていただ	いたし/Depr	き委員会の採決は6月10日に

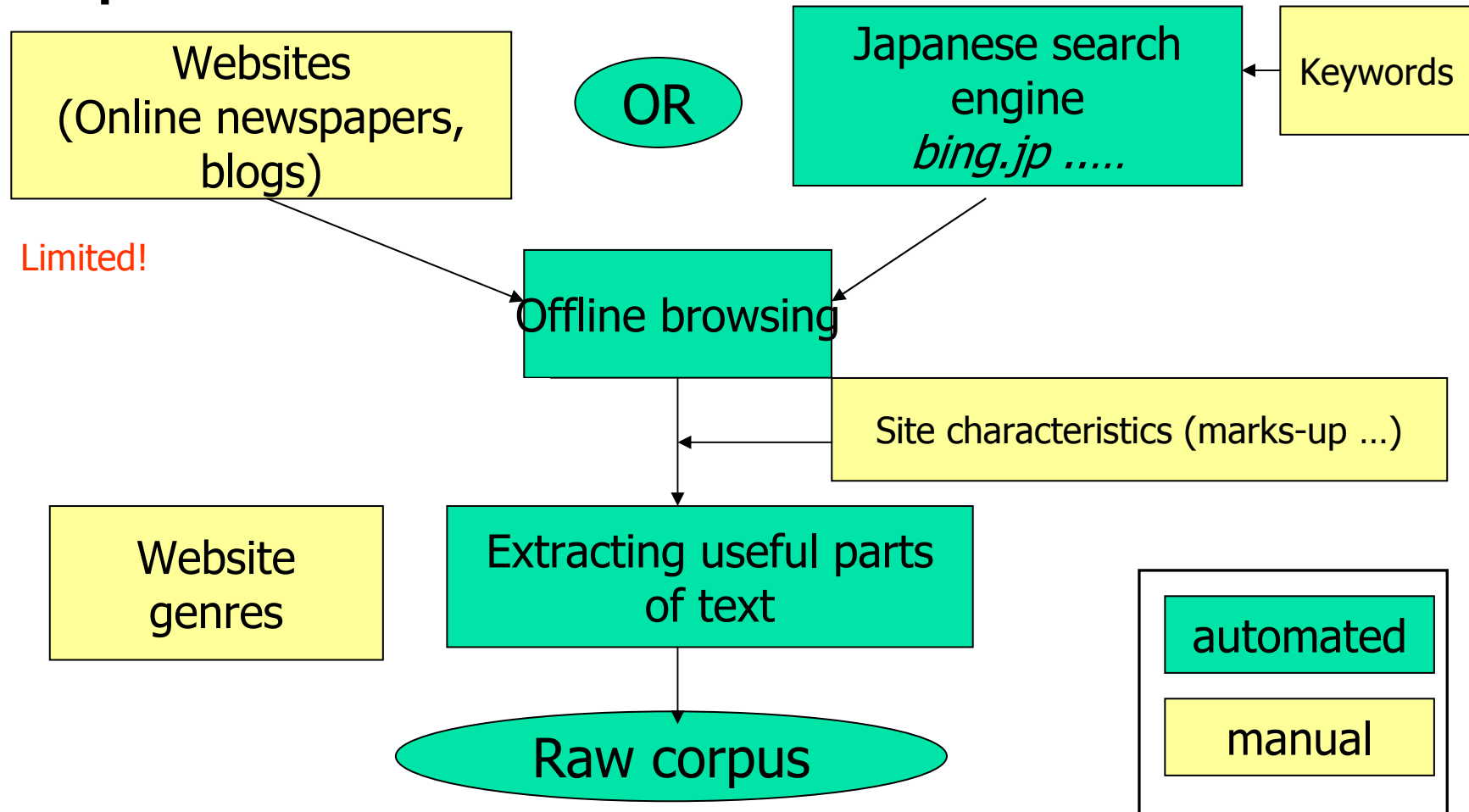
NooJ KWic with annotation

GRAM = honorifique 87/87

- **Corpora and tools**
 - Segmentation problem
 - Selected corpora and tools
 - BCCWJ (corpus)
 - SAGACE (corpus + tool)
 - NooJ (tool)
 - **Open search on the web**

Using corpora to find relevant examples of Japanese honorific forms in order to generate exercises for intermediate learners of Japanese

■ Open search on the Web





**Using corpora to find relevant examples of Japanese honorific forms
in order to generate exercises for intermediate learners of Japanese**

- Research context
- Why do we need corpus analysis ?
- Corpora and tools
- **Adopted methodology for the use of corpora**
- Future research perspectives



Using corpora to find relevant examples of Japanese honorific forms in order to generate exercises for intermediate learners of Japanese

Adopted methodology

- Identifying characteristic forms from other studies on honorifics
- Extracting these forms from Japanese corpus incorporated in parsing tools (Sagace, BCCWJ)
- Identifying the most frequently used honorific forms
- Doing further research on the Web (updates)
- Identifying text-genders and contexts of uttering
- Creating annotated corpus from the raw elements obtained
- Formalizing the rules of generation
- Validating of the generated sentences by native Japanese speakers



**Using corpora to find relevant examples of Japanese honorific forms
in order to generate exercises for intermediate learners of Japanese**

- Research context
- Why do we need corpus analysis ?
- Corpora and tools
- Adopted methodology for the use of corpora
- **Future research perspectives**



Using corpora to find relevant examples of Japanese honorific forms in order to generate exercises for intermediate learners of Japanese

Future research perspective

- Analysing other texts to increase the corpus size (identification and selection of contexts =a significant time-consuming manual phase)
- Classifying contexts of use and defining the data representation for context of the utterance (relational Database / ontology?) (Current prototyping in Prolog)
- Implementing the generated sentences in a e-learning environment
- Application to other linguistic contexts



Using corpora to find relevant examples of Japanese honorific forms in order to generate exercises for intermediate learners of Japanese

Bibliography

- **Blin, R. 2001.** *Traitement automatique du japonais et études linguistiques.* Histoire Epistémologie Langage vol 23 n° 1 p. 33-48.
- **Kishimoto, H. 2010.** *Possessor Raising and Two Types of Honorification,* Nanzan Linguistics No.6, p 1-12.
- **Maekawa, K. 2007** "Design of a Balanced Corpus of Contemporary Written Japanese." *Proceedings of Symposium on Large-Scale Knowledge Resources (LKR2007),* pp.55-58, 2007:03.
- **Rastier, F. (dir.) , 1995** *L'analyse thématique des données textuelles,* Paris : Didier, 1995 p. 223-249.
(Texte légèrement remanié.)
- **Schikowski, R. 2008** *Skript zum Strukturkurs Japanisch.* LMU Munich, Institute of General and Typological Linguistics, winter term 2007/08
- **Shimamori, R. 2001** *Grammaire japonaise systématique,* volume II - Edition Jean Maisonneuve Paris 12ème édition, 2001



Using corpora to find relevant examples of Japanese honorific forms in order to generate exercises for intermediate learners of Japanese

Bibliography

- **Siegel M. 2000.** *Japanese Honorification in an HPSG Framework*, Proceedings of the 14th Pacific Asia Conference on Language, Information and Computation. p. 289-300.
- **Sugimura, R. 1986.** *Japanese honorifics and Situation Semantics* International Conference on Computational Linguistics COLONG. p. 507-510.
- **Terrya, K. 2007.** *Interpersonal grammar of Japanese* in A Systemic functional grammar of Japanese, Vol 2, Ch 4, p135-205.
- **Trusty, A. / Truong, K 2011** *Augmenting the Web for Second Language Vocabulary Learning* Conference on Human Factors in Computing System (CH)., pp 3179-3188
- **Wlodarczyk, A. 1996.** *Politesse et Personne – Le japonais face aux langues occidentales*. Editions L'harmattan.
- **Wlodarczyk, A. 2007.** *Towards a Unified Treatment of Linguistic - Person and Respect - Identification* Japanese Linguistics – European Chapter Tokyo.
- **Zock M. / Lapalme G. 2010** *A Generic Tool for Creating and Using Multilingual Phrasebooks*. Natural Language Processing and Cognitive Science.



**Using corpora to find relevant examples of Japanese honorific forms
in order to generate exercises for intermediate learners of Japanese**

Merci vielmol fers zuhere

