One and half years of Japanese community

maho@openoffice.org

中田真秀

(NAKATA, Maho)

Purpose of this presentation

Japanese is spoken almost in Japan. We use over 10,000 letters, mainly brought from China and Korea (Kanji), and also developed its own letters (Hiragana and Katakana). After the Meiji period, we also began to use Alphabets.

This indicates treating Japanese demands a lot of computer resouces, and its history is also unique.

From ja native lang project, let non-native people know Japanese and Japanese culture better for good communication, and hope this encourages other native language projects, and for developers, we can share good experiences of treating such a complicated language.

Agenda

- About the speaker
- Two topics :
    - *Brief introduction of Japanese and Japan*
    - *Our results in 1.5 years*
        - *QA in 1.1.0, 1.1.1*
        - *Site management*
        - *Books, articles*
        - *CD-ROM*
        - *Development*

About the speaker

- 中田真秀 (NAKATA, Maho)
- Contributions for OpenOffice.org
  - Japanese native lang project lead since its beginning (2003/Jan/17~)
  - Porting and providing packages for FreeBSD
    - http://porting.openoffice.org/freebsd/
- Other activities in free software community
  - FreeBSD ports committer since 2002/Oct/18
    - Maintaining ~30 packages (w/o OO.o)
    - Total 625 commits for ports tree
- Post doctoral fellow at Tokyo University
  - quantum chemistry solves Schrödinger equation, without experiments. I'm very happy to have a presentation here at Erwin Schrödinger-Zntrum!

Brief introduction of Japanese (1)

- In China, about B.C. 1300, people began to use letters (there was no letters in Japan at that time).

- Consists of several parts : typically 偏 (hen ; real part) and 旁 (tsukuri; abstract part)

- 猫 (cat) consists of two parts; `犭' and `苗'. is pronounced as `mieu'  and 犭' means animal; cat is a kind of animal crying like `mieu'.

animal   Mieu

犭   苗

猫 cat

好 (like) ; woman likes her child.

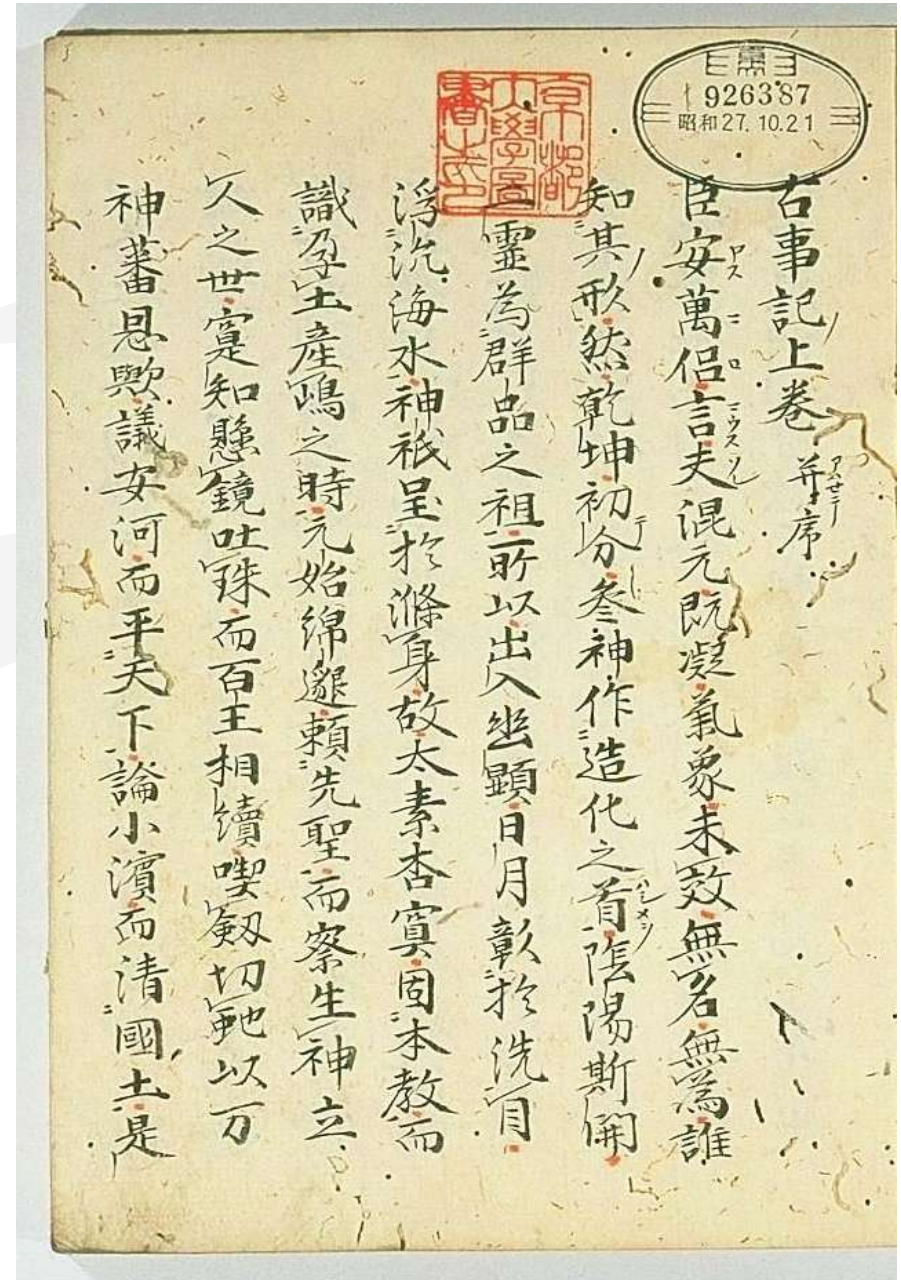Woman 好 Child

好 Like

Brief introduction of Japanese (2)

- We didn't not used letters until about 5 A.D.
- Letters are brought from China and Korea (漢字 ; Kanji) and read as Japanese as phonogram (later also used as ideogram)
- Written in *vertically*
- Famous writings
  - 古事記 (Kojiki; history of Japan, A.D. 712)
  - 万葉集 (Manyôu syu, A.D. 730?)
- With a writing brush

Copyright 2001, Kyoto University Library

古事記

Brief introduction of Japanese (3)

- We have developed our own letters : ひらがな (Hiragana) and カタカナ (Katakana)
- Hiragana and Katakana were made from a part of or collapsed Kanji and phonogram
- the shape of letters has been unified almost at the beginning of 11 A.D.

From Kanji To Hiragana
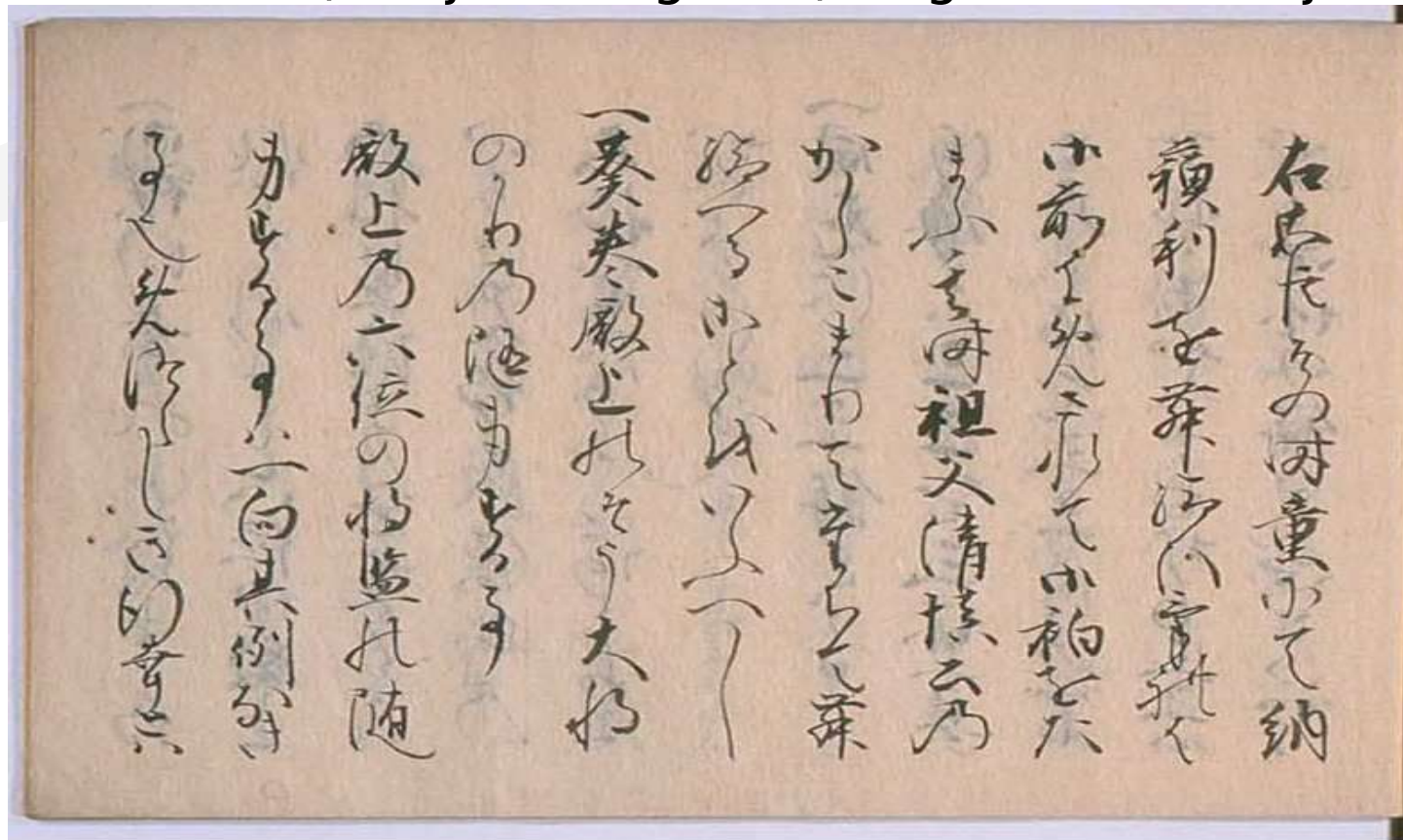
From Kanji To Katakana

阿→ア
伊→イ
宇→ウ

Time evolution

Brief introduction of Japanese (4)

- Hiragana expresses one's emotions; written in beautiful curves, and always the subject of art.
- Kanji and Katakana had been used in formal documents.

源氏物語 (Genji-monogatari) hirgana and Kanji

Copyright 2002, Kyoto University Library

Brief Introduction of Japanese (5)

- We use three kind of letters： ひらがな (Hiragana)，かたかな (Katakana) and 漢字 (Kanji)
  - 今日はよい天気ですね。
- Hiragana/Katakana are phonogram, 47 letters , everything can be written in Hiragana or Katakana (like Hangul)

ひらがな　　　　　　　　　　　　　カタカナ

んわらやまはなたさかあ　　　ンワラヤマハナタサカア
　り　みひにちしきい　　　　　　リ　ミヒニチシキイ
　うるゆむふぬつすくう　　　　　ウルユムフヌツスクウ
　れ　めへねてせけえ　　　　　　レ　メヘネテセケエ
　をろよもほのとそこお　　　　　ヲロヨモホノトソコオ

Complete table of Hiragana and Katakana, and also an example of vertical (table) and horizontal (caption) writing...

Brief Introduction of Japanese (5)

- Written in both horizontally (usually left to right) and vertically

Horizontal writing

おはようございます。今日はよい天気です。

Examples of Vertical Writing

今日は平成十六年九月二十二日です。

皆さんこんにちは。OpenOffice.orgで日本語を表示した例

Brief introduction of Japanese (6)

- Over 125,000,000 people uses mainly in Japan (8th spoken language)
- Characteristics of Japanese grammar
  - *verb is always located at the end of the sentence*
    - *SOV (subject+object+verb)*
    - *SCV (subject+complement+verb)*
  - No space between words
    - We distinguish words by intonation (spoken), and Kanji or punctuation chars (written)
    - 単語と単語の間に空白は入りません。発音では抑揚を手がかりに表記では漢字や句読点を手がかりに単語を区別します。
- Many homonyms
  - きかい、機械 (machine)、機会 (opportunity)、奇怪 (strange)

Brief introduction of Japanese (7)

- Many meanings in one sentence. Same pronunciation, but different meaning.
  - うらにわには、にわにわとりがいる。
    - In the backyard, there are two cooks.
  - うらには、にわ、にわとりがいる。
    - In back of the house, there are two cooks.
  - うらに、はにわにわとりがいる。
    - In the backyard, there is (are) Haniwa-cook(s).

Brief Introduction of Japan (1)

- Political situation:
  - *Quite strong bureaucratism*
  - *Conservative*
  - *Still Japanese feels ambiguous oppression or fear by `Okami' : boss, government, Emperor and U.S.A. etc.*
- People:
  - *Do one's work very thoroughly*
  - *Tame, but oppressive character*
- Communication method:
  - *Do not say yes or no clearly*
  - *Do not say one's opinion clearly*
  - *Slow in decision making*
  - *Prefer to stay anonymous*

Brief Introduction of Japan (2)

- Strong affinity for other cultures and customs
  - *Very flexible for .foreign words are used as Japanese immediately, using Katakana.*
    - コンピューター *(computer)* ベルリン *(Berlin)*
  - *You can enjoy almost every cuisine around the world :)*
- Culture
  - *Pop: animation, manga, game, karaoke*
  - *Movies, Haiku, Sushi, Bushi-do, Ukiyo-e*
- Other minor languages spoken in Japan:
  - *Aynu (Hokkaido)*
  - *Uchina-guchi (Okinawa)*

Japanese language processing (1)

- Coding system by JIS (Japanese Industrial Standards)
  - Character code : JIS X 0208-1997 (6,355 kanji)
- Unicode 2.1
  - 20,902 letters; also including Chinese, Korean and Vietnamese
  - No compatibilities between Unicode and JIS X 0208-1997; requires large translation table...
- Several Encoding systems are available:
  - EUC-JP (ISO 2022 complaint)
  - JIS (or ISO-2022-JP)
  - Shift JIS
    - MS has been using it so influencing; not ISO complaint

Japanese language processing (2)

- We must handle 10,000 chars, then input method (IME) is also a problem.
- Hiragana's Alphabetical notation : Romaji
  - *Aiueo, kakikukeko*（あいうえお、かきくけこ）
- Alphabet->Romaji->Hiragana->Kanji+hiragana
  - Watashi no Namae Ha Nakata desu. (romaji)
  - わたしのなかたです。 (hirangana onjy)
  - 私の名前は中田です。 (kanji and hiragana)

Japanese language processing (3)

- Grammar is difficult to treat, so we employ heuristics to translate. typical IME has:
  - *Several megabytes of dictionary (Hiragana to Kanji)*
  - *Grammar analysis engine (heuristic, a lot of know-hows)*
    - *Find a separation between words (maybe several possibilities ; user selects from the candidates)*
    - *Smart IME recognise the meaning of the sentence (still may have some possibilities)*

- Proprietary IMEs (smart)
  - ATOK, Vje, MS IME (Windows), Kotoeri (Mac OS X)
- Free IMEs (not smart ; for UNIX)
  - Wnn, Canna and SKK

Japanese language processing (4)

- Availability of truetype fonts
  - *There are no high quality free (in a sense, MIT or lesser license) fonts*
  - *Many proprietary fonts : problems incompatible between them*
- Availability of bitmap fonts
  - *There are free bitmap fonts came with X11*

Break

Next : problems we solved in/and introduction of ther results of Japanese OO.o community

QA and localized builds (1)

- 1.1 from 1.0 : major milestone for ja people now supports forbidden chars, many i18n enhancements.

- Serious problem before OO.o 1.1.0 release

  - *Wrong PDF export #iZ 15444*
    - *Regression in 1.1Beta2 and 1.1rc. #iZ 14788 and #iZ 11579 were fixed.*

- OO.o 1.1.1 release : #iZ 27865

  - *In OO.o 1.1.1, there are still wrong PDF export, and fix is available, however, cws are not integrated in RC.*
  - *We fixed by ourselves, and QA'ed.*
  - *We released our own binary package!*

    build policy, QA protocole; what can we do?

QA and localized builds (2)

- Build policy of Hamburg Release Engineering team:
  - *Currently, 10 major langs and Windows/Linux/Solaris are bulid by Hamburg Release Engineering team.*
  - *Who should do the build?*
  - *Who controll which cws should integrated in mws between RC?*
  - *What is the reason-d'être of native-lang?*
    - *Though QA is usually done by native-langs, however, we sometimes cannot `stop' the release even if it has regressions or bug can be workarounded/not integrated in mws.*

QA and Localized builds (3)

- For 1.1.2, we won the codename of `Hakone' by vote!
- For 1.1.3, ja hosted the vote and Berlin won the codename

SourceCast and Internet Explorer (1)

- SourceCast is a project managing software for OpenOffice.org web site.
- I18n'ed SourceCast still seems to have a problem with Internet Explorer (IE)...
  - *Header always contains*
    <meta http-equiv="Content-Type" content="text/html; charset=UTF-8" /> , though
  - *Still IE doesn't work this header, for first access, default settings of IE regard this page as encoded by Shift-JIS.*
  - *Mozilla, FireFox, Opera and Safari doesn't have such problem.*
  - *Currently we have no idea to fix this...*
- Workaround : IZ #28839, IZ#30665 (not fixed yet)

SourceCast and Internet Explorer (2)

Workaround : IZ #28839, IZ#30665

Site management problem

- Historical reason and small flexibility in ja.oo.o site, we also run several websites:
- http://oooug.jp
  - *Backup, BBS, Bug tracking system, etc.*
- http://openoffice.s16.xrea.com:8080/pukiwiki/pukiwiki.php?FrontPage
  - *Puki Wiki site*
- http://openoffice-docj.sourceforge.jp/
  - *Documentation, translation*
- http://sourceforge.jp/projects/waooo/
  - *Upload community builds*
- http://search.luky.org/oo/
  - *Archiving mailing list (we still use outside of ja.oo.o resource)*
- Apparently, this complicates managing...

Newsletter

- We publish a news letter : 3 letters/month
- http://ja.openoffice.org/marketing/newsletter/
    - *Very good source, very well managed.*
- Main readers are casual users

Translation

- We have been translating texts around the world : Danny's power tools, etc. Discussions are held at translation@ja.openoffice.org.
- Ja native lang talked with Young Song (globalization project lead of SUN), and decided that OO.o ja community and SUN do translation in cooperation
  - *Ja community had never done translating UI and help, however several persons have already raised their hands.*
  - *How we cooperate is still not well established. it has just started now.*

Contribution from ja community

- Better Japanese support for DXF. #iZ 9391
- Add signature information XML-signature form by W3C (SCOP)
    - *http://www.3bit.co.jp/scop/ Security Component of OpenOffice.org Plugin*
    - *Based on OpenPGP*
- Japanese translation of OpenOffice.org developer's guide
    - *http://www.opendt.org/pukiwiki/pukiwiki.php?Translation*

Development (2)

- Mac OS X localization :
  - *Better detection of locale so that we can use Japanese*
- Improvement in Calc, and file handling
  - IZ #3875, 27955
  - iZ #12313
- FreeBSD porting
  - *1.1.x : almost (90%) done, ready for upcoming 1.1.3*
  - *2.x : builds and works: milestone 41*
  - *# of patches are now ~10. last year we have >120.*
  - *Now we have porting status page:*
    - *http://porting.openoffice.org/freebsd/*

Women ML had discontinued

- Women ML started at : 2003/Jan/19th
    - women@ ja.openoffice.org
- Discontinued : 2004/July/26th
- Main topic was:
    - *Discuss marketing for women*
- Very small traffic : total 285 mails (before suggesting of removal : 156 mails)
- Good example of how to add/remove mailing lists. Should be careful as far as possible, and Proposer should have strong motivation.

High quality Japanese fonts (1)

- For free desktop environment, availability of public high quality Japanese fonts is extremely important.
- Also several demands for Japanese fonts from non-native community.
- Information-technology Promotion Agency (IPA) bought a commercial full set of Japanese fonts and distributed under quite generous license.

IPA 明朝

愛のある豊かな日本語表現

High quality Japanese fonts (2)

- Still we have font problem:
  - Windows use proprietary fonts : MS 明朝 /MS ゴシック
  - Commercial Linux distributions have same fonts but different name
  - Free Linux distributions have low quality fonts.
  - StarOffice have its own fonts (HG Mincho Light J)
    - *All fonts are not compatible*
    - *Induces so called Toufu, all the Kanji, hiragana are displayed as   or destroys layout.*

- Rigid font treatment:
  - Use some virtual fonts : carries only metrics.
  - Possible choices of low/high quality fonts
    - If you want a high quality font, just buy it.
  - PostScript and TeX employ such kind of virtualise mechanism. See #iZ 15611, 20370 for details.

Published Books (1)

- "OpenOffice.org Official User's Guide" was put on market on April 23. The authors of this book are ja.oo.o members. That's why we call it Official. ISBN4-83991-141-X, 441 pages, also planning to publish at ja.oo.o in html format.

- IDG Japan, Inc. publishes "OpenOffice.org FAQ for Linux Users.", ISBN4-87280-212-8, It has 1 CD with J2RE, MySQL Connector and other useful tools.
  http://www.idg.co.jp/campaign/mook5-1/index.html

CD-ROM project

- From jaooo-cdrom project CD-ROM project released several CD-ROMs

Version 1.1.1

407 CD-ROM
was distributed
at KO+F 2003
(Ver 1.1.0)

62 CD-ROMs
was distributed
at KO+F 2003
(Ver 1.1.0)

Attended events

- http://ja.openoffice.org/marketing/event/index.html
- 2004
  - *OpenSource Conference (Tokyo; 2004/Sep/4)*
  - *GODo2004 (Hokkaido ; 2004/Jul/24)*
  - *Mozilla Party 5.0 (Tokyo ; 2004/Apr/18)*
- 2003
  - *Kansai Opensource+Freeware 2003 (Osaka, 2003/Oct/31~Nov/1)*
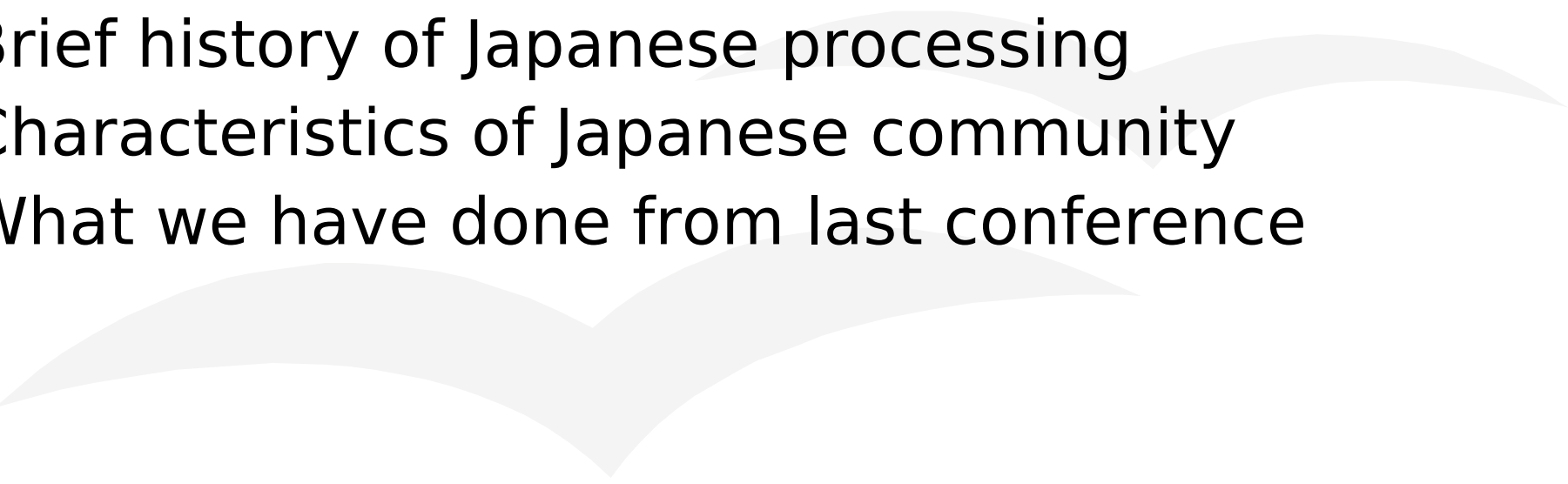- Have a booth, sessions, distributes CD-ROMs, and so on.

Contribution with Good-day inc.

- Manuals
  - *http://desktop.good-day.net/ooo11/man.php*
  - *Introduction guide are distributed with official OO.o package*
  - *Open Manual : for beginners.*
  - *Open Text : for seminar use.*
- Source code reading seminar
  - *Reading 1.1.2 and 2.0 codes.*
- Templates (mainly from OO.o extra)
  - *http://desktop.good-day.net/ooo11/tmp.php*
- FAQ
  - *http://desktop.good-day.net/ooo11/faq.php*
- Applications
  - *http://desktop.good-day.net/ooo11/api.php*

Summary

- Introduction of Japanese
- Brief history of Japanese
- Brief history of Japanese processing
- Characteristics of Japanese community
- What we have done from last conference

Acknowledgement

Thank you very much!

We are supported by

OO.o community,

Japanese OO.o community, and

Free Software Initiative of Japan

(FSIJ)