

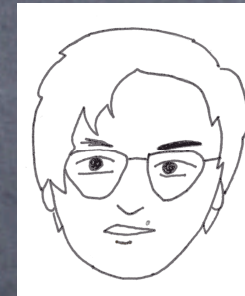
自然な訳文を生成する 翻訳システムの開発

むしゃ

武舎 広幸

マーリンアームズ（株）

河村 政雄



2003年7月12日
未踏15キックオフ・セミナー
京都リサーチパーク

昨年度の未踏事業

自分と同レベルの訳文を出力する 翻訳システムの探求

→ 機械翻訳エンジンの開発

利用者のことではなく……

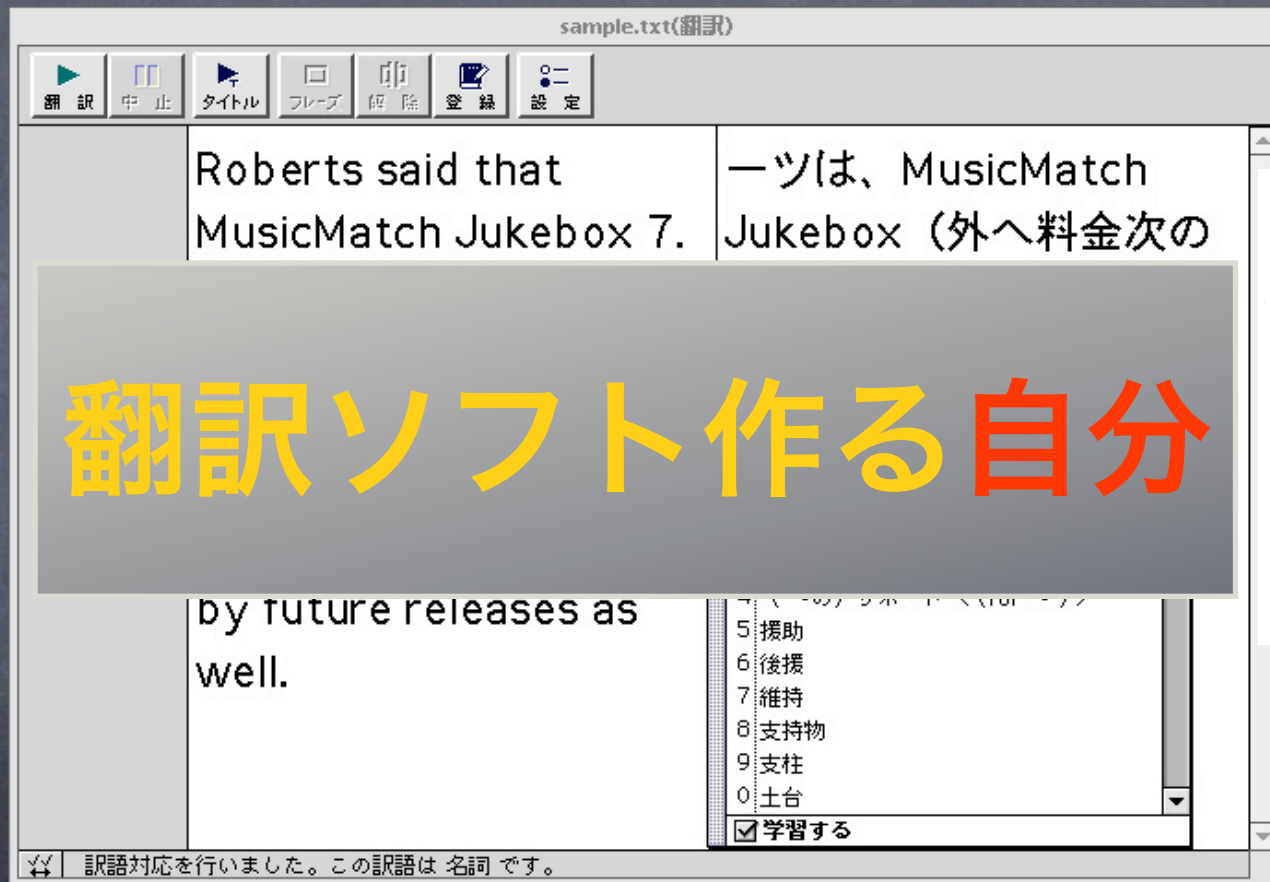
開発者（武舎）のこと

自分と同レベルの訳文を出力する 翻訳システムの探求

翻訳をする自分



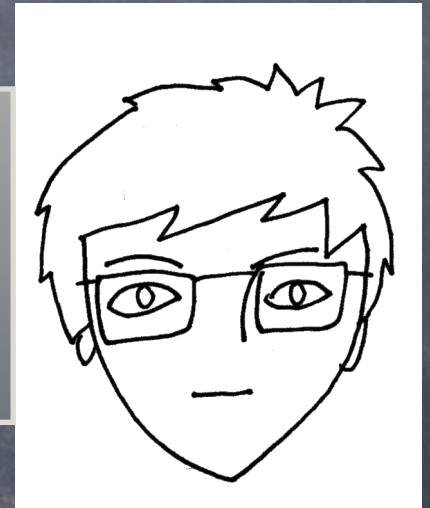
自分と同レベルの訳文を出力する 翻訳システムの探求



自分と同レベルの訳文を出力する 翻訳システムの探求

翻訳を教える自分

監修 武舎広幸



自分と同レベルの訳文を出力する

翻訳システムの探求



自分と同レベルの訳文を出力する 翻訳システムの探求

自分が目指してきた訳

- ・ 自然な訳文
- ・ 読みやすくわかりやすい訳文
- ・ 原文の情報をもれなく伝える
- ・ 原文の世界と等価な世界を作り上げる

十九世紀なかばのオーストラリア。大陸北東部クイーンズランドの開拓が、徐々に徐々に北へと進み、ようやく東海岸のなかほどに達した、その最前線の村でのこと。ある日、村はずれの放牧場で子供たちが遊んでいると、異様なものが現われた。子供は三人。つぎのあたった木綿の服を着た姉妹と、その従兄弟でズボンつりに半ズボンの少年——いずれも裸足であたりを駆け回る、ちょっとのことでは動じない農家の子である。

ONE DAY IN the middle of the nineteenth century, when settlement in Queensland had advanced little more than halfway up the coast, three children were playing at the edge of a paddock when they saw something extraordinary. They were two little girls in patched gingham and a boy, their cousin, in short pants and braces, all three barefooted farm children not easily scared.

(David Malouf, *Remembering Babylon*)

自分と同レベルの訳文を出力する 翻訳システムの探求

自分が目指してきた訳

- ・ 自然な訳文
- ・ 読みやすくわかりやすい訳文

→ どうしてよいのか検討がつかない

別のアプローチ

- ・ おかしな（機械翻訳臭い！）訳を少なくする
- ・ ひとつひとつ表現に対応していく
- ・ 対応できるように枠組みを拡大する

機械翻訳臭い訳文とは 例 1

今後数週間で発表されるMusicMatch Jukebox 7.5は、将来のリリースによっても維持される iPodサポートを得る。

MusicMatch Jukebox 7.5 -- due out sometime in the next several weeks -- will have iPod support that should be maintained by future releases as well.

機械翻訳臭い訳文とは 例 1

今後数週間で発表されるMusicMatch Jukebox 7.5は、将来のリリースによっても維持される iPodサポートを得る。

MusicMatch Jukebox 7.5 -- due out sometime in the next several weeks -- will have iPod support that should be maintained by future releases as well.

その単語だけをみて訳している

機械翻訳臭い訳文とは 例2

コンピュータ上でブートする時、MusicMatchは
最新バージョンの7.2にアップグレードしたいか
どうか確認してくる。

When you boot up the computer, MusicMatch
asks if you want to upgrade to the newest
version, 7.2.

品詞に縛られている

これに対して人間の翻訳は

これに対して人間の翻訳は

数週間のうちにリリースされる予定の

MusicMatch Jukebox 7.5では、iPodがサ

ポートされる予定であり、以後のリリース

ス **まわりを見て柔軟な対応ができる**

コンピュータのブート時に、MusicMatch

は新しいバージョン、すなわち7.2にアップ

グレードするかどうか尋ねてくる。

このシステムが目指すもの

機械翻訳臭くない訳文の生成



まわりを見て柔軟な対応ができる

どのような翻訳ができるようになったか

目指したもの

機械翻訳臭くない訳文の生成

まわりを見て柔軟な対応ができる

高速辞書引きソフトを超えるもの

どのような翻訳ができるようになったか

MusicMatchとiPodに関してひとつ奇妙なのはデバイスが動作するソフトウェアのバージョンである。iPodにはMusicMatchソフトウェアのバージョン7.1が付属している。コンピュータのブート時に、MusicMatchは新しいバージョン、すなわち7.2にアップグレードするかどうか尋ねてくる。ほとんどの人はいつも最新のソフトウェアを使うためにアップグレードをしたいと思うだろうが、そうするとiPodを利用できなくなってしまう。

この問題は次のマイナーバージョンアップで修正されるはずである。MusicMatch担当者のJennifer Robertsは（今後数週間のうちに出る予定の）MusicMatch Jukebox 7.5がiPodをサポートするだろう、そしてそれは今後のリリースによっても変わらないはずであると語った。



どのような翻訳ができるようになったか

One strange thing about MusicMatch and the iPod is the software versions the device will work with. The iPod comes with version 7.1 of the MusicMatch software. When you boot up the computer, MusicMatch asks if you want to upgrade to the newest version, 7.2 -- while most people would do the upgrade to stay current with the latest software, you won't be able to use your iPod if you do.

This problem should be remedied with a new point release. MusicMatch spokesperson Jennifer Roberts said that MusicMatch Jukebox 7.5 -- due out sometime in the next several weeks -- would have iPod support that should be maintained by future releases as well.



他の「システム」との比較

システム A

(今後数週間のうちに出る予定の) MusicMatch Jukebox 7.5がiPodをサポートするだろう, そしてそれは今後のリリースによっても変わらないはずであると語った。

システム B

MusicMatchジュークボックス7.5(次の数週間でいつか予定されている)が将来のリリースによって同様に維持されるべきiPod支援を得るだろうと言った。

システム C

今後数週間で発表されるMusicMatch Jukebox 7.5は、将来のリリースによっても維持されるiPodサポートを得るという。

This problem should be remedied with a new point release. MusicMatch spokesperson Jennifer Roberts said that MusicMatch Jukebox 7.5 -- due out sometime in the next several weeks -- would have iPod support that should be maintained by future releases as well.

他の「システム」との比較

システム A

(今後数週間のうちに出る予定の) MusicMatch Jukebox 7.5がiPodをサポートするだろう, そしてそれは今後のリリースによっても変わらないはずであると語った。

システム B

MusicMatchジュークボックス7.5(次の数週間でいつか予定されている)が将来のリリースによって同様に維持されるべきiPod支援を得るだろうと言った。

システム C

今後数週間で発表されるMusicMatch Jukebox 7.5は、将来のリリースによっても維持されるiPodサポートを得るという。

システム D

数週間以内にリリース予定のバージョン7.5からはiPodがきちんとサポートされるとのことだ。

他の「システム」との比較

システムA 未踏

（今後数週間のうちに出る予定の）MusicMatch Jukebox 7.5がiPodをサポートするだろう、そしてそれは今後のリリースによっても変わらないはずであると語った。

システムB 機械翻訳ソフトX

MusicMatchジュークボックス7.5(次の数週間でいつか予定されている)が将来のリリースによって同様に維持されるべきiPod支援を得るだろうと言った。

システムC ニュースサイトXの訳文（翻訳ソフト使用？）

今後数週間で発表されるMusicMatch Jukebox 7.5は、将来のリリースによっても維持されるiPodサポートを得るという。

システムD 人間翻訳（武舎）

数週間以内にリリース予定のバージョン7.5からはiPodがきちんとサポートされるとのことだ。

他の「システム」との比較

システムA 未踏

(今後数週間のうちに出る予定の) MusicMatch Jukebox 7.5が iPodをサポートするだろう、そしてそれは今後のリリースによっても変わらないはずであると語った。

システムB 機械翻訳ソフトX

MusicMatchジュークボックス7.5(次の数週間でいつか予定されている)が将来のリリースによって同様に維持されるべきiPod支援を得るだろうと言った。

MusicMatch Jukebox 7.5 -- due out sometime in the next several weeks -- would have iPod support that should be maintained by future releases as well.

他の「システム」との比較

システムA 未踏

(今後数週間のうちに出る予定の) MusicMatch Jukebox 7.5が iPodをサポートするだろう、そしてそれは今後のリリースによっても変わらないはずであると語った。

システムB 機械翻訳ソフトX

MusicMatchジュークボックス7.5(次の数週間でいつか予定されている)が将来のリリースによって同様に維持されるべきiPod支援を得るだろうと言った。

MusicMatch Jukebox 7.5 -- due out sometime in the next several weeks -- would have iPod support that should be maintained by future releases as well.

他の「システム」との比較

システムA 未踏

(今後数週間のうちに出る予定の) MusicMatch Jukebox 7.5が iPodをサポートするだろう、そしてそれは今後のリリースによっても変わらないはずであると語った。

チューニングしている

システムB 機械翻訳ソフトX

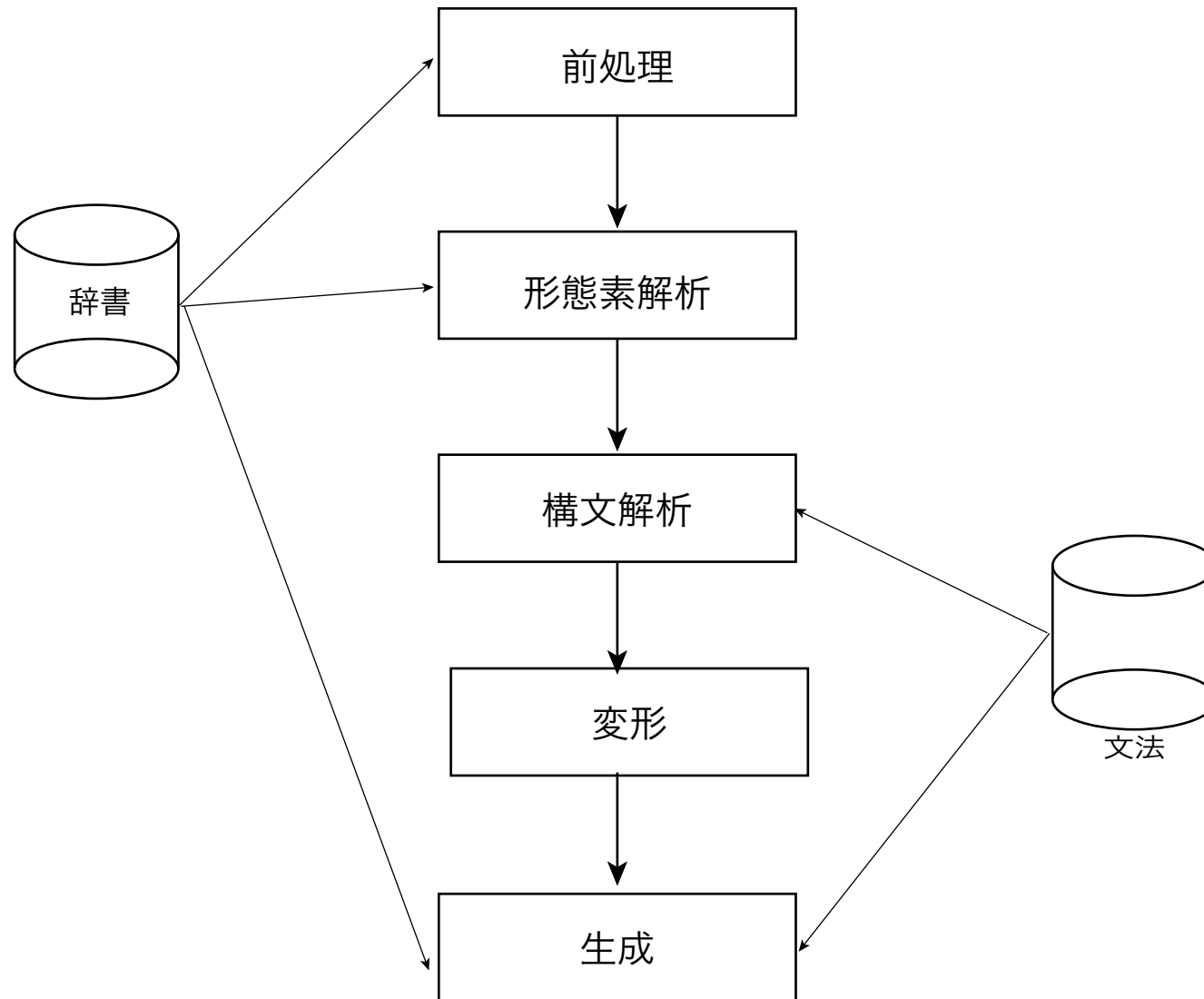
MusicMatchジュークボックス7.5(次の数週間でいつか予定されている)が将来のリリースによって同様に維持されるべきiPod支援を得るだろうと

チューニングしたくてもできない！

MusicMatch Jukebox 7.5 -- due out sometime in the next several weeks -- would have iPod support that should be maintained by future releases as well.

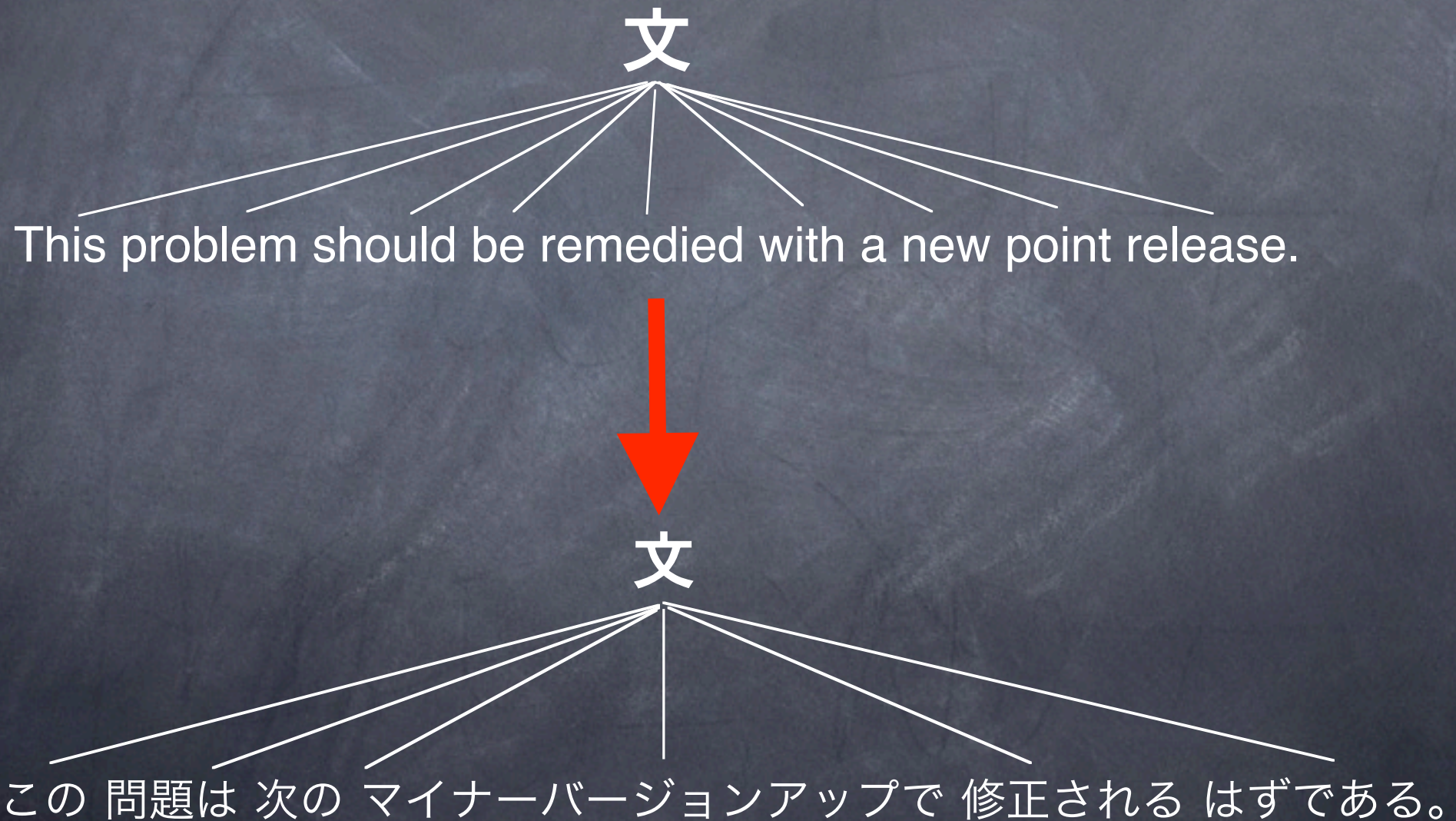
昨年度作成した
システムの概要

システムの概要



システムの基本原理

翻訳の過程を木構造の変形として捉える



システムの基本原理 — 辞書引き

文

This problem should be remedied with a new point release.



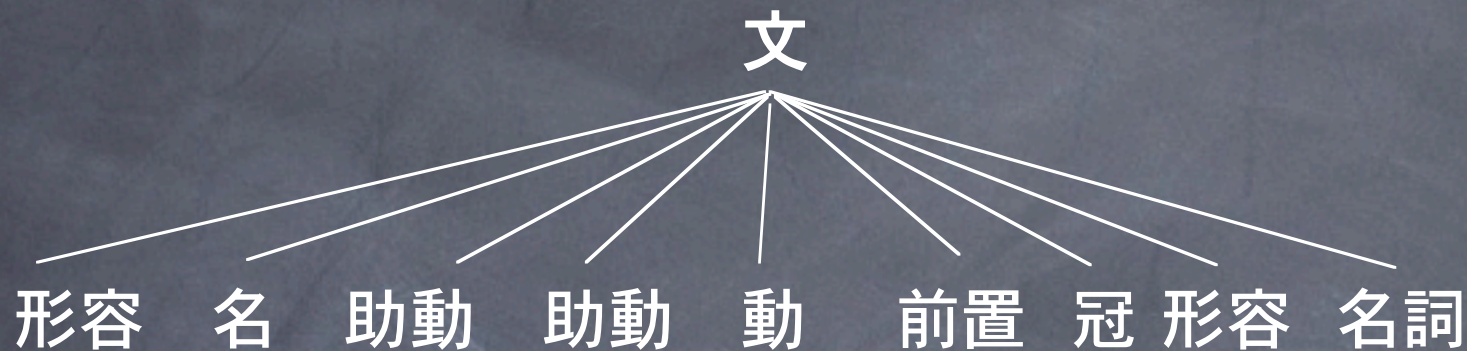
形態素解析
(辞書引き)

文

形容 名 助動 助動 動 前置 冠 形容 名詞

This problem should be remedied with a new point release.

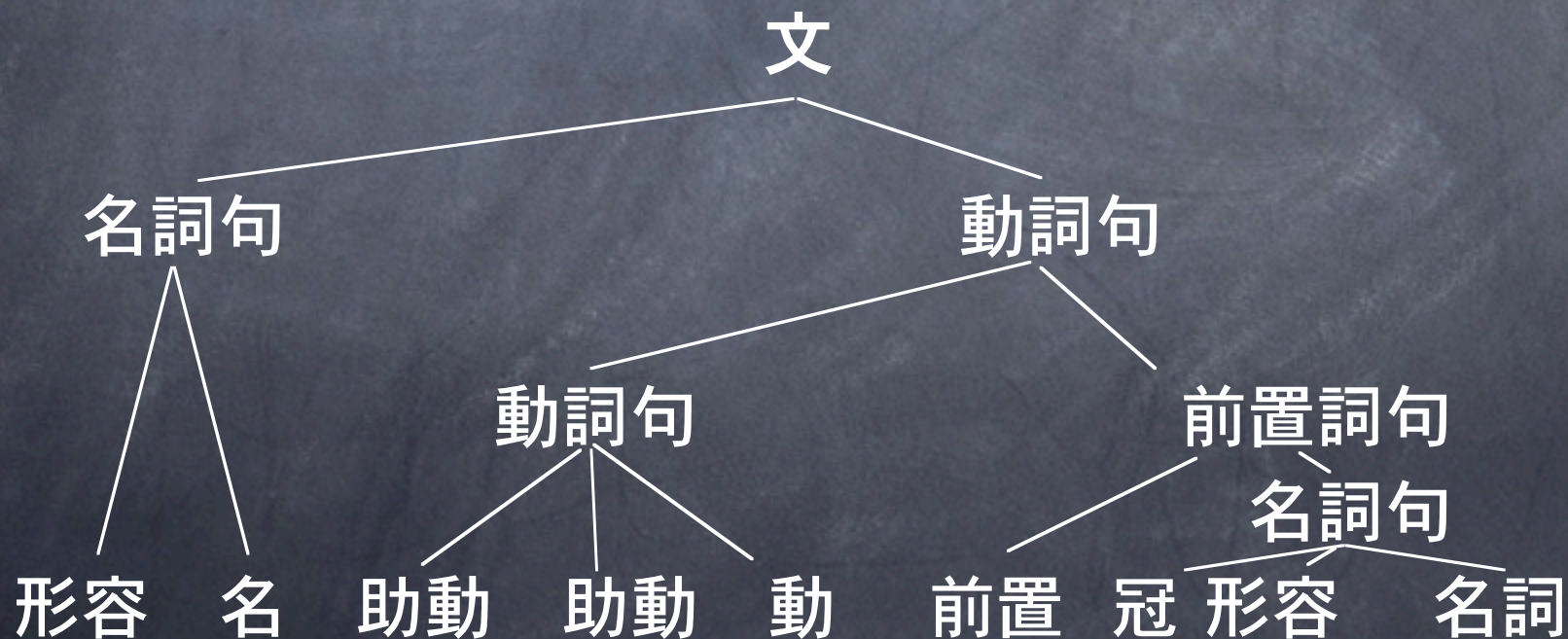
システムの基本原理 — 構文解析



This problem should be remedied with a new point release.

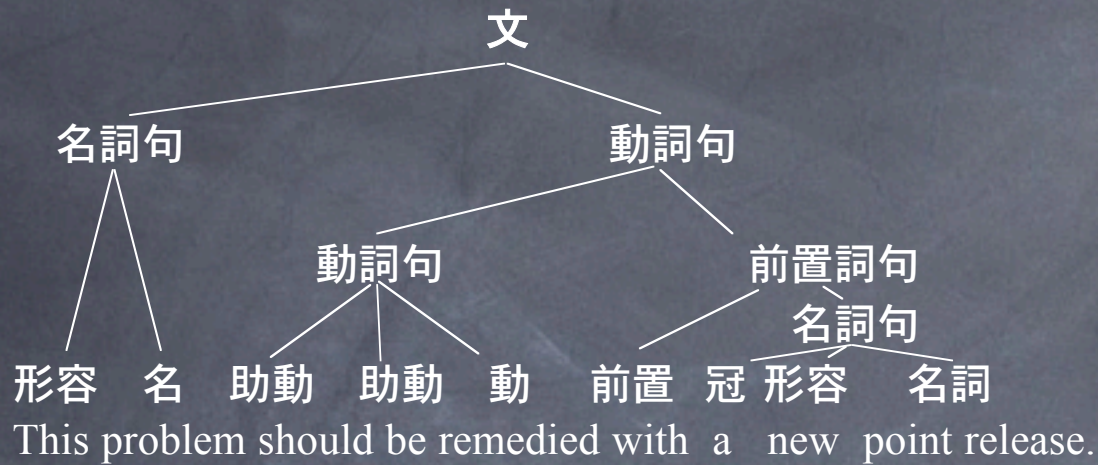


構文解析 (文法規則)

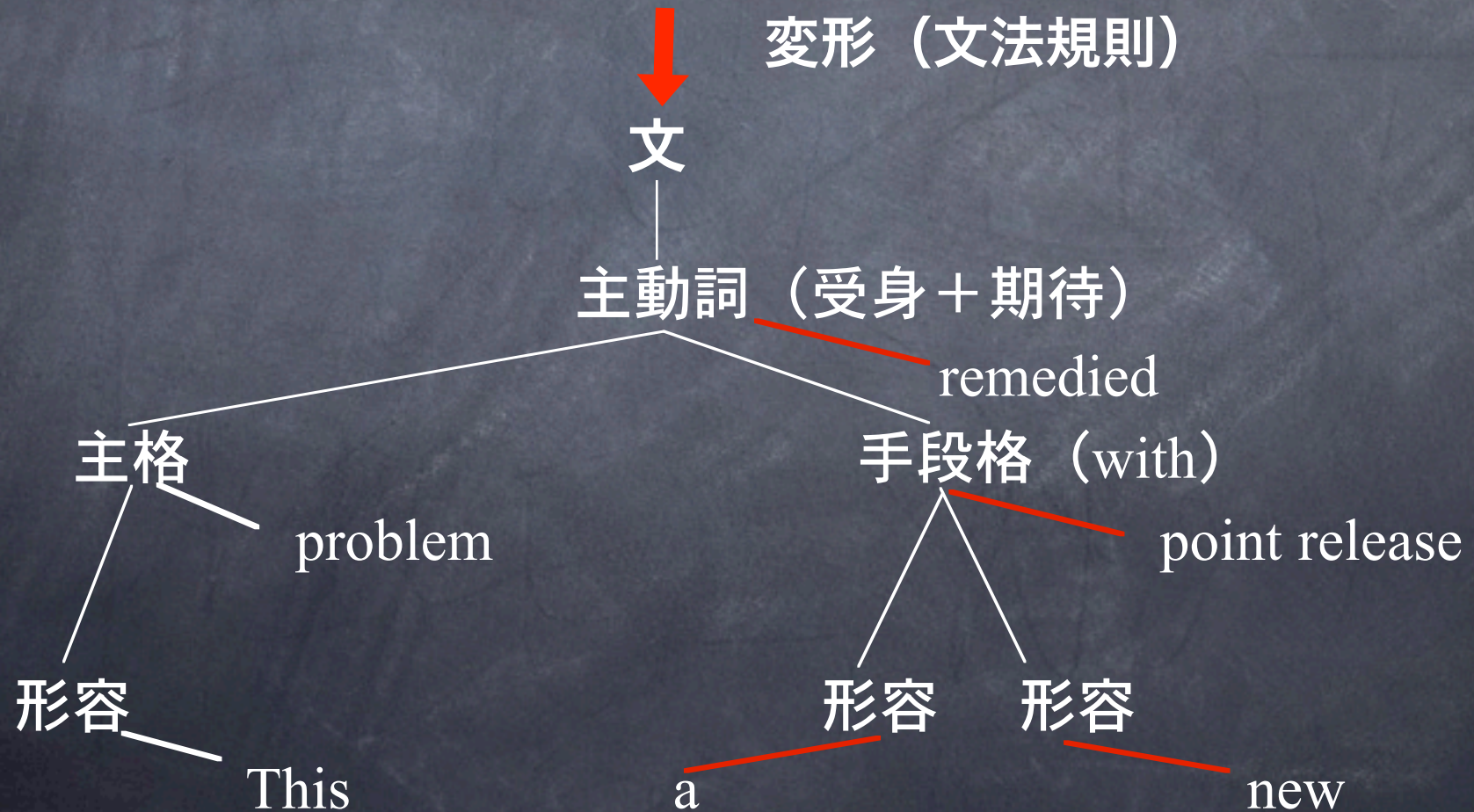


This problem should be remedied with a new point release.

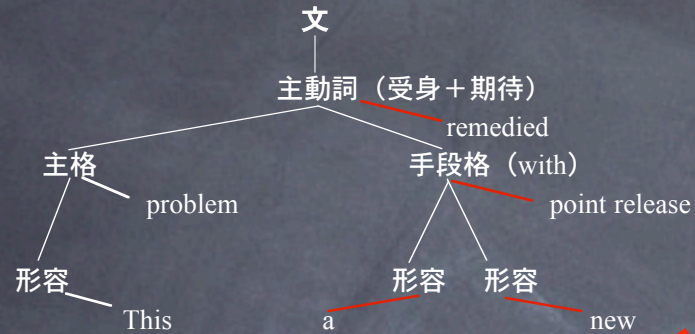
システムの基本原理 — 変形



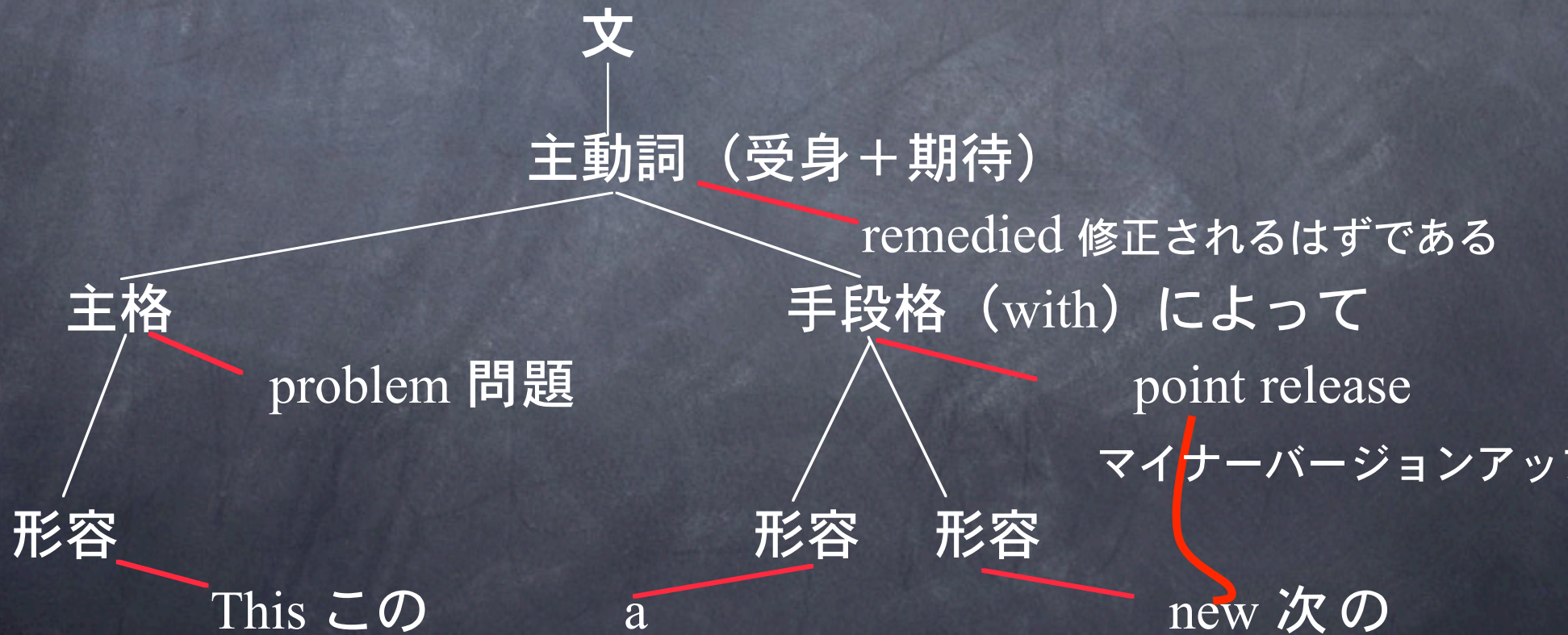
変形 (文法規則)



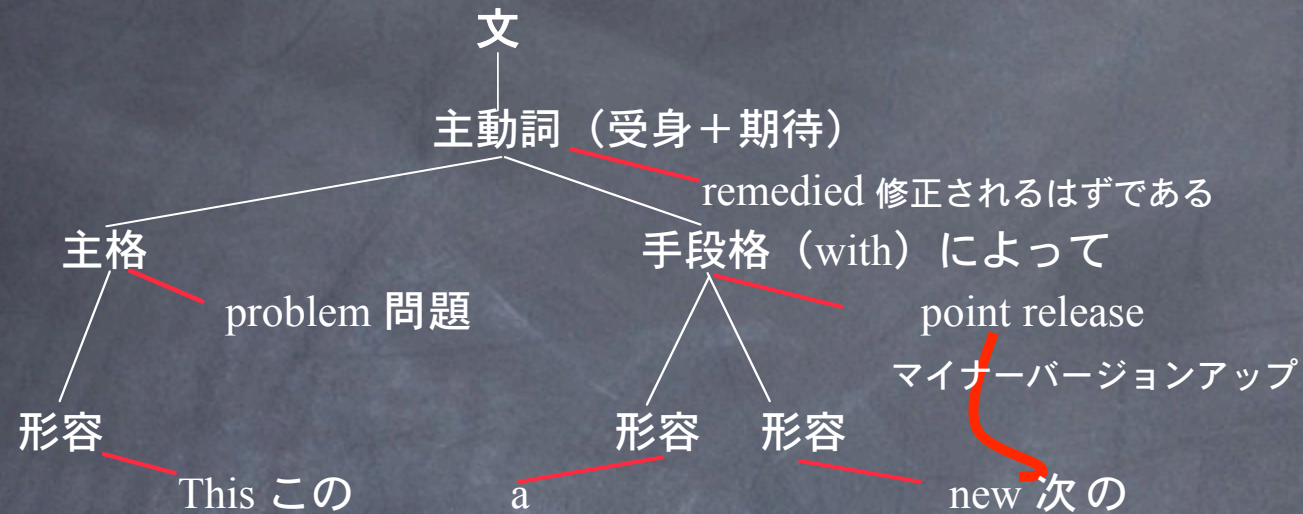
システムの基本原理 — 生成



生成 (辞書)



システムの基本原理 — 出力



出力

文

この 問題は 次の マイナーバージョンアップで 修正される はずである。

特徴

まわりの状況を見て

柔軟な変形操作が行える

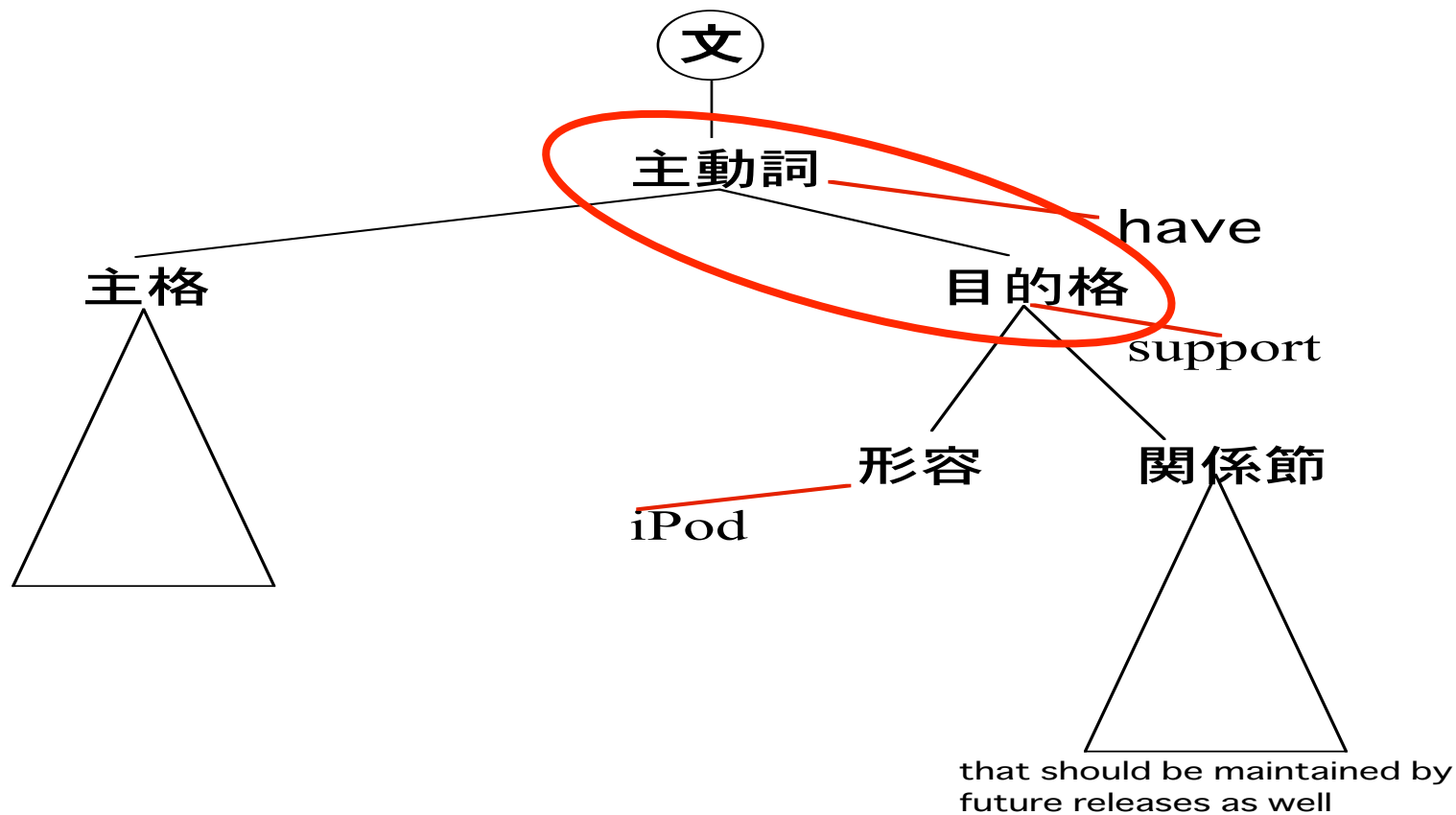
特徴

MusicMatch Jukebox 7.5 -- due out sometime in the next several weeks -- would **have iPod support** that should be maintained by future releases as well.

（今後数週間のうちに出る予定の）MusicMatch Jukebox 7.5が iPodを**サポートする**だろう、そしてそれは今後のリリースによっても変わらないはずである。

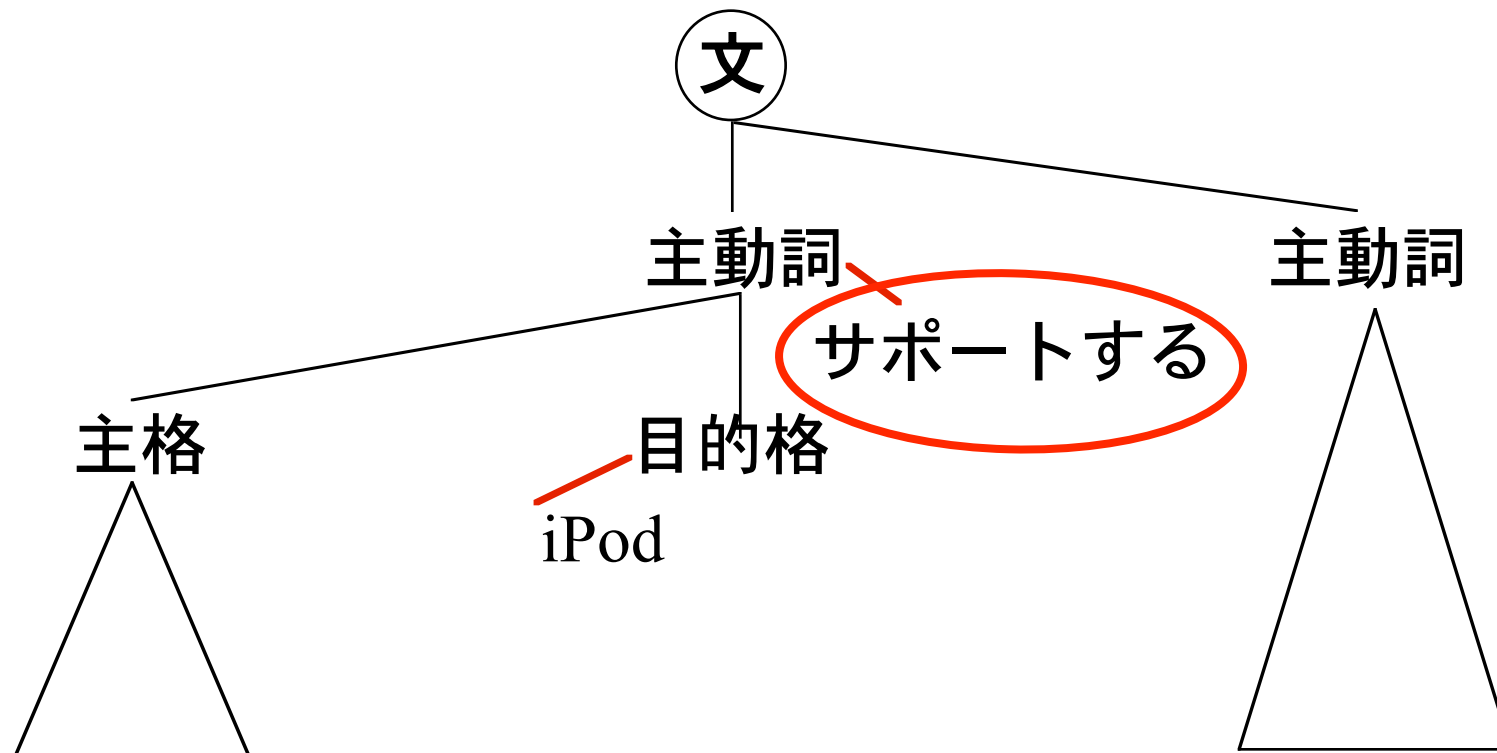
特徴

MusicMatch Jukebox 7.5 -- due out sometime in the next several weeks -- would **have iPod support** that should be maintained by future releases as well.



特徴

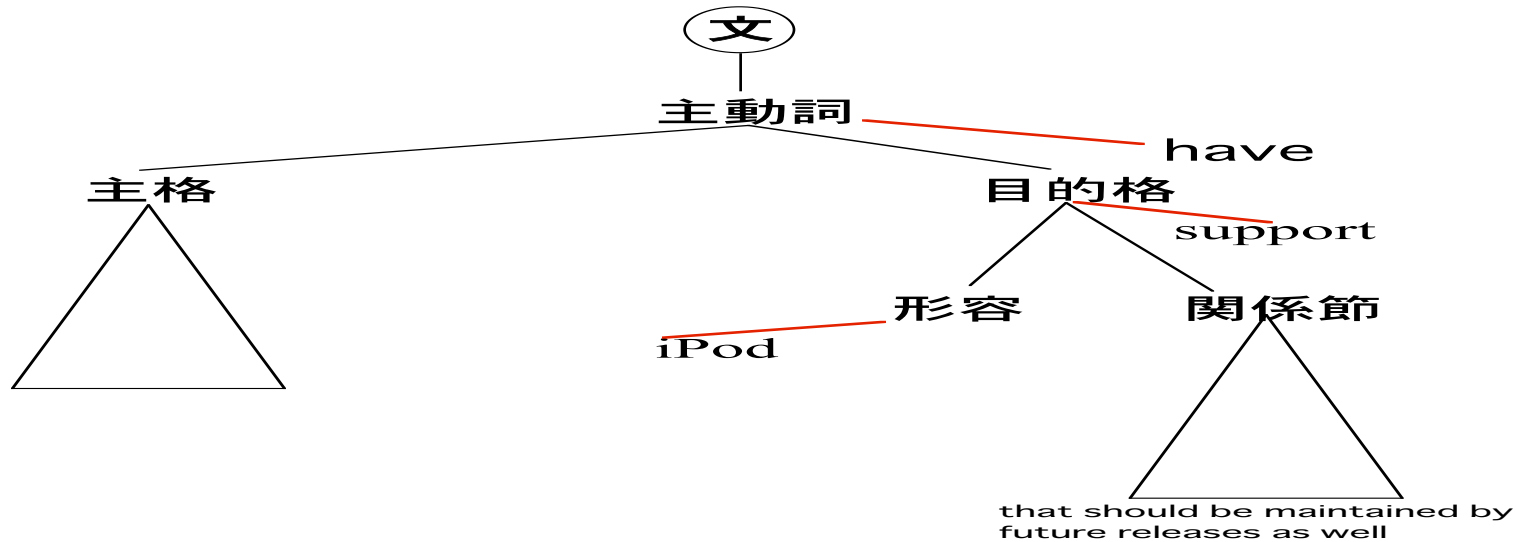
(今後数週間のうちに出る予定の) MusicMatch Jukebox 7.5がiPodをサポートするだろう, そしてそれは今後のリリースによっても変わらないはずである。



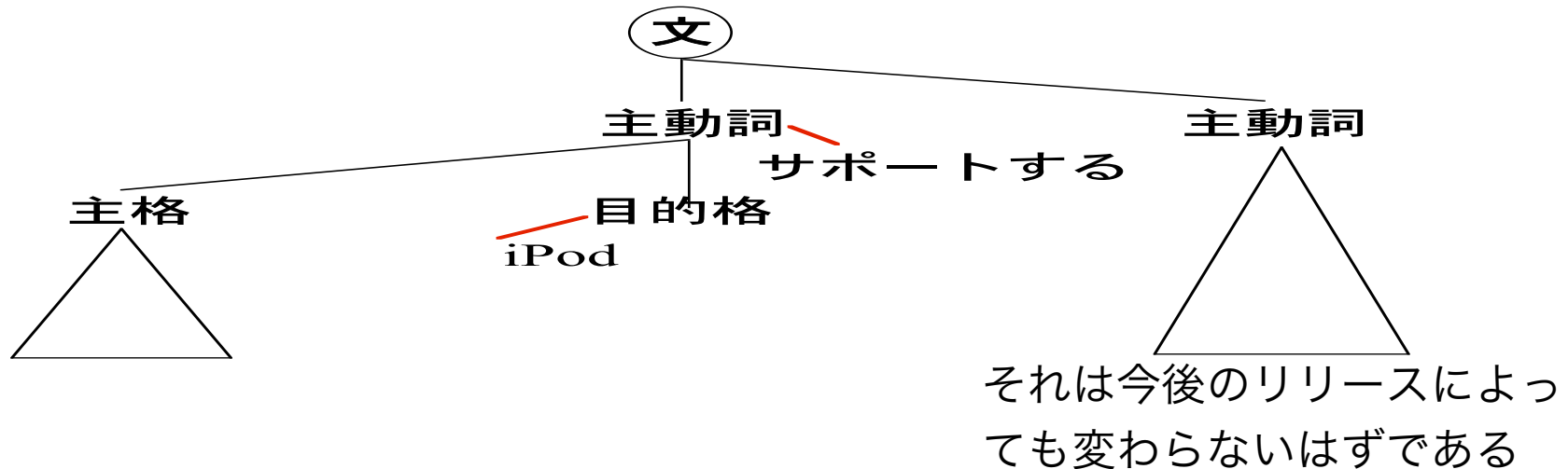
それは今後のリリースによっても変わらないはずである

特徴

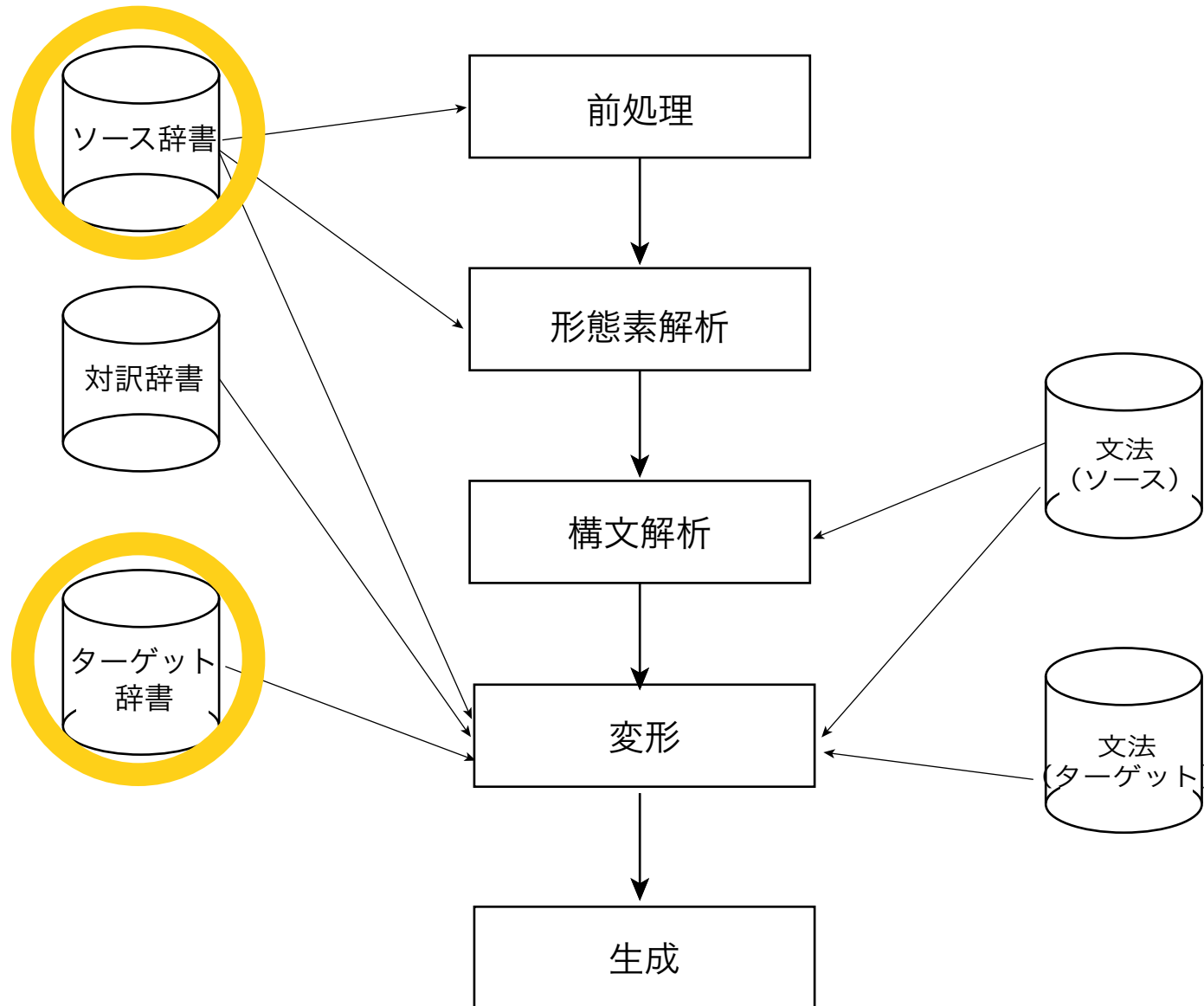
haveを起点に条件を記述



変形操作を行う



辞書の構成



ターゲット言語辞書 — 例文

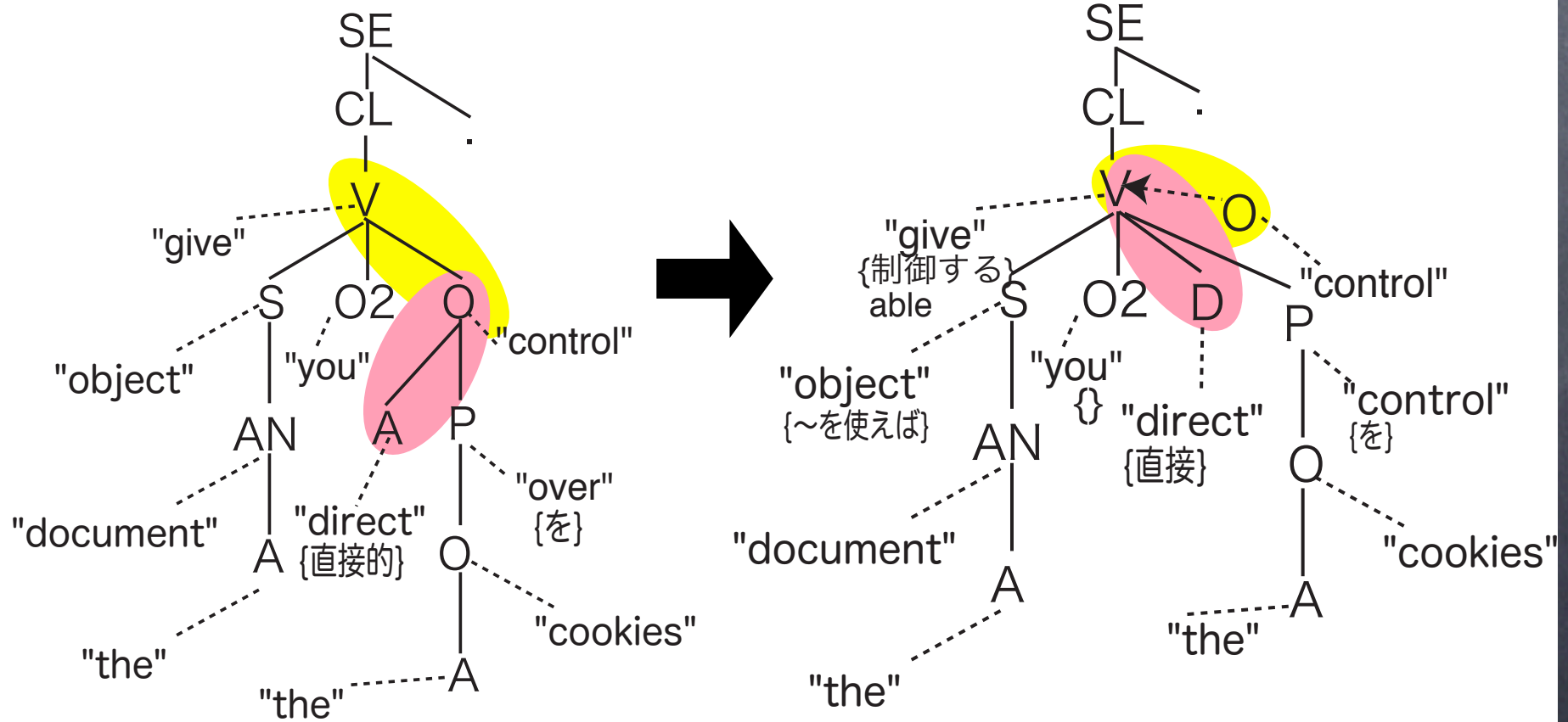
The document object gives you direct control over the cookies.

documentオブジェクトは、**あなたに**クッキーの**直接的制御を与える**。

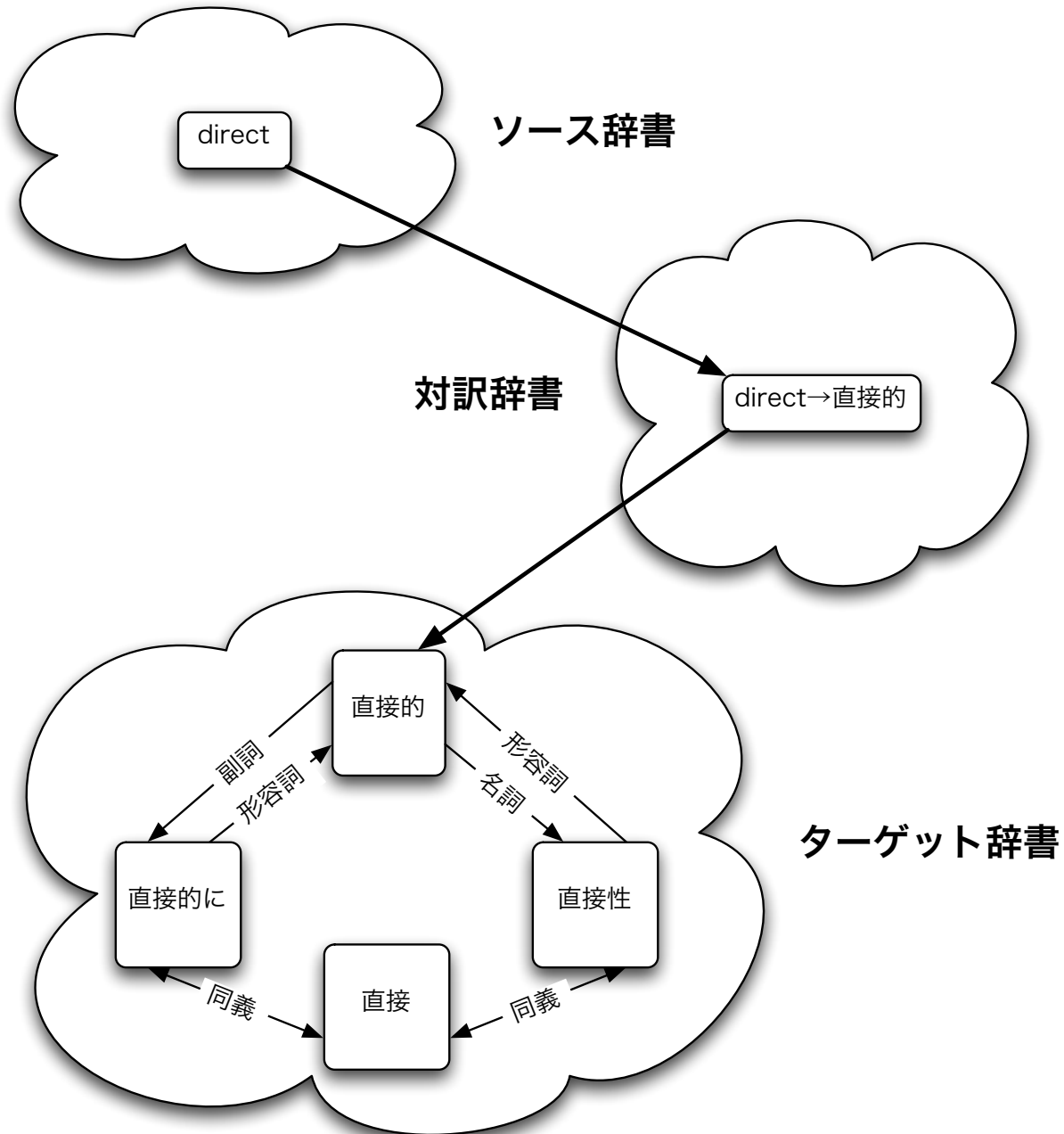
documentオブジェクトを使えばクッキーを**直接制御できる**。

ターゲット言語辞書

The document object gives you direct control over the cookies.



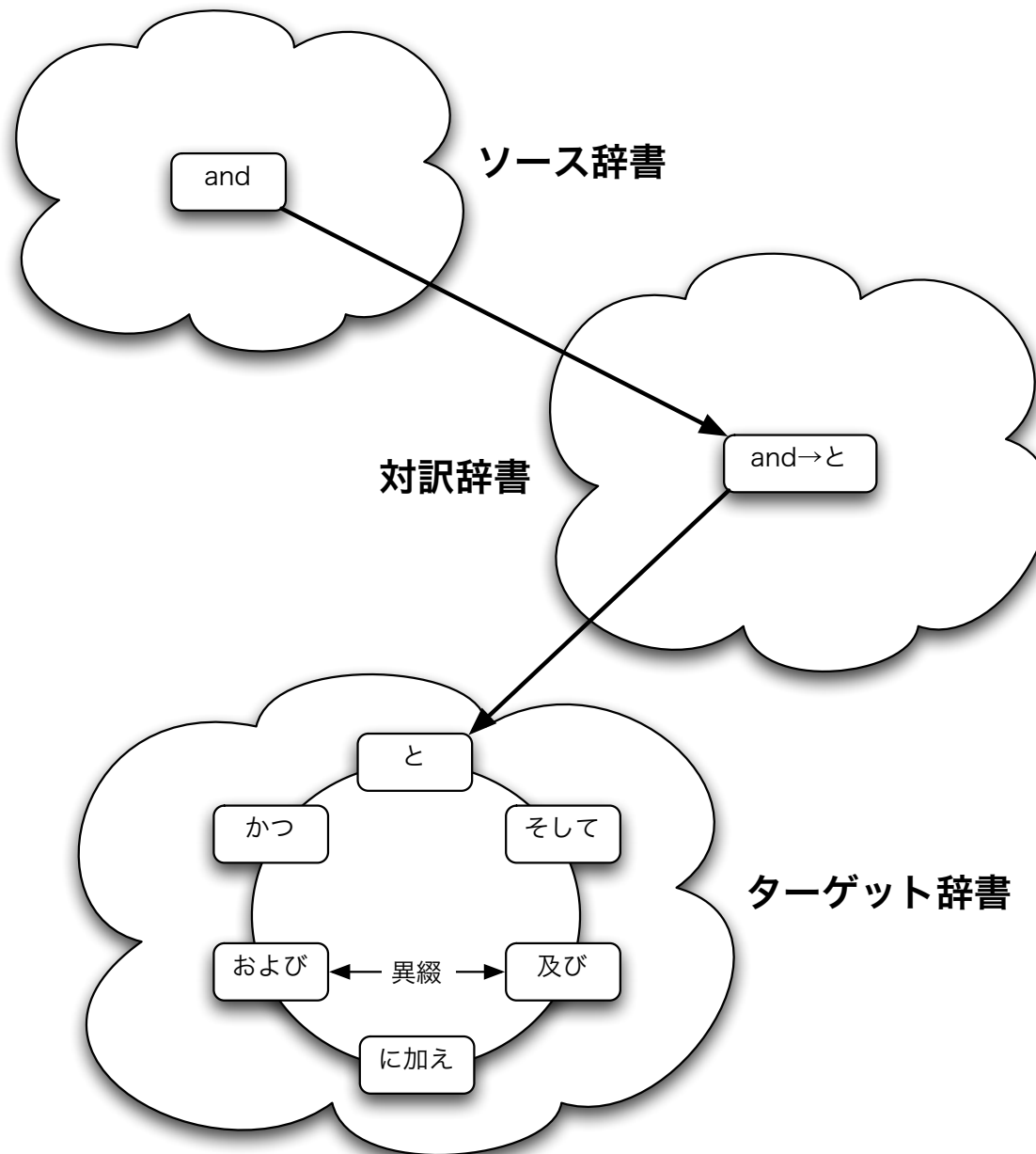
ターゲット言語辞書



ターゲット言語辞書

and
1 と
2 および
3 そして
4 及び
5 かつ
6 且つ

ターゲット言語辞書



ターゲット辞書

and	
1 と	▼
2 および	▶ 及び
3 そして	
4 かつ	▶
5 に加え	



優先する訳語の指定

表現は

- ☒ フォーマルな単語を優先（購入する、販売する）
- ☐ 一般的な単語を優先（買う、売る）

接続詞は

- ☒ 平仮名語を優先（および、かつ）
- ☐ 漢字混じりの語を優先（及び、且つ）

カタカナ語の最後の長音を

- ☒ 付ける（コンピューター、ユーザー）
- ☐ 付けない（コンピュータ、ユーザ）

カタカナ語と漢字語の訳がある場合

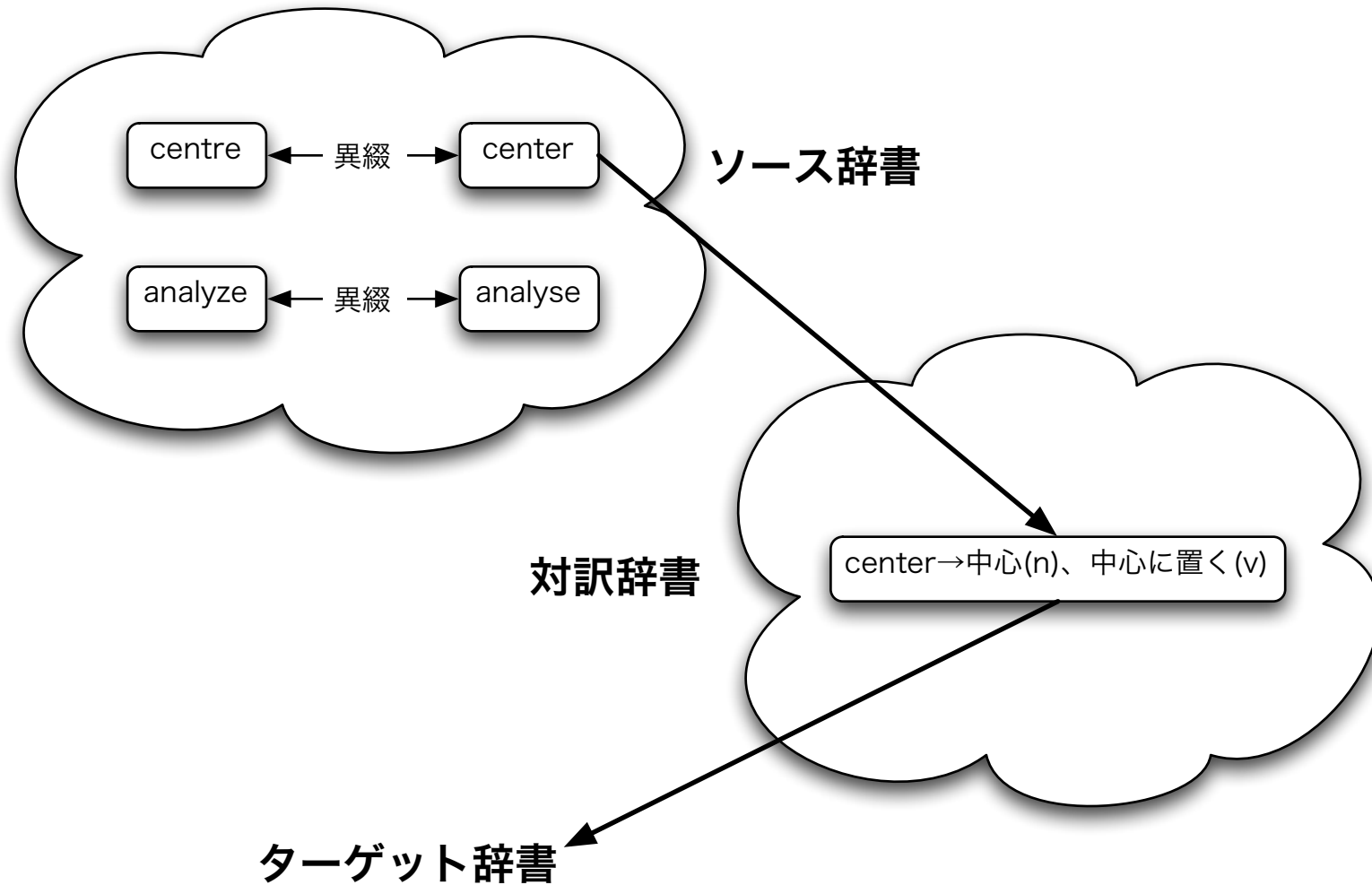
- ☒ 一般的な語を優先（利用者、ウィンドウ）
- ☐ カタカナ語を優先（ユーザー、ウィンドウ）
- ☐ 漢字語を優先（利用者、窓）

数字は

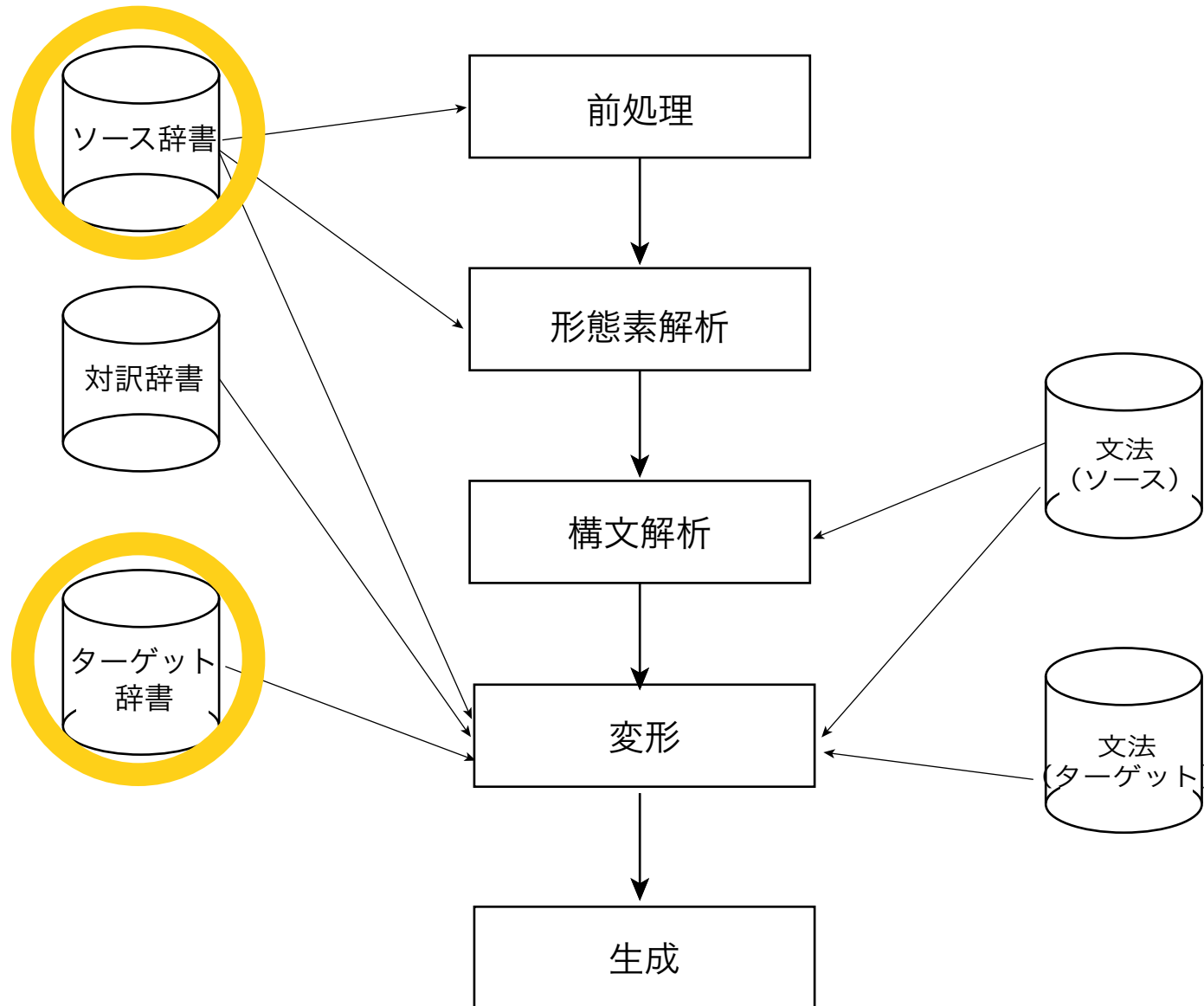
キャンセル

OK

ソース言語辞書



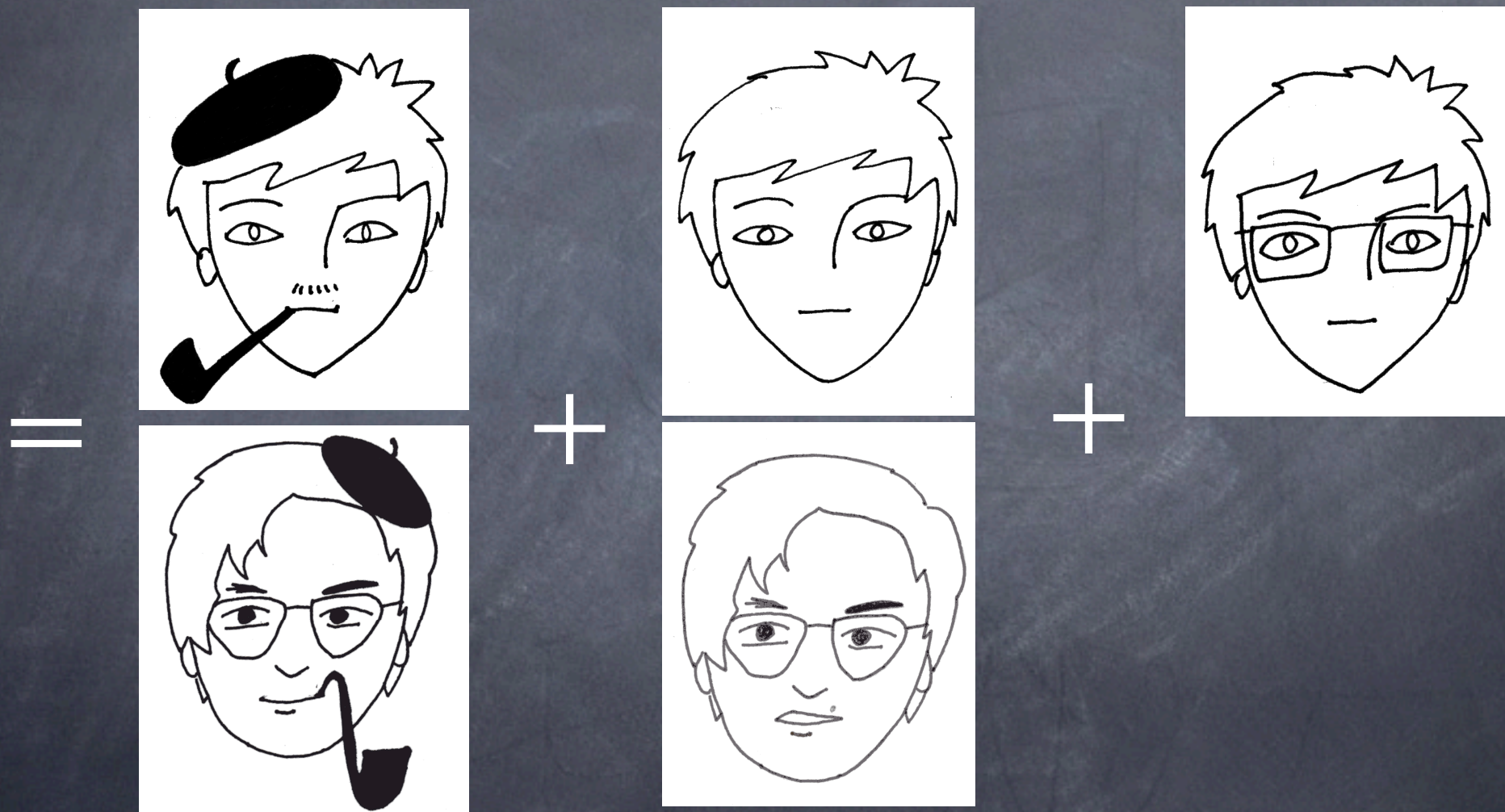
辞書の構成



今年度の開発

今年度

自然な訳文を生成する 翻訳システムの開発



翻訳をする我々 ソフトを作る我々 翻訳を教える我々

今年度の開発

- ・ プロ翻訳者向けのツール
- ・ 辞書データの構築

ビジネスショーにて

「何がおかしいんですか？」

システムC

ニュースサイトXの訳文

今後数週間で発表されるMusicMatch Jukebox 7.5は、将来のう。

質が問われなくなっている

役に立たない？

今年度の開発

- ・ ~~プロ翻訳者向けのツール~~
- ・ ~~辞書データの構築~~

ビジネスショーにて

「いつ製品になるんですか？」

「辞書を作らなくてはいけないので、
2年後ぐらいには製品化したいんですが……」

辞書の構築は大変

今年度の開発

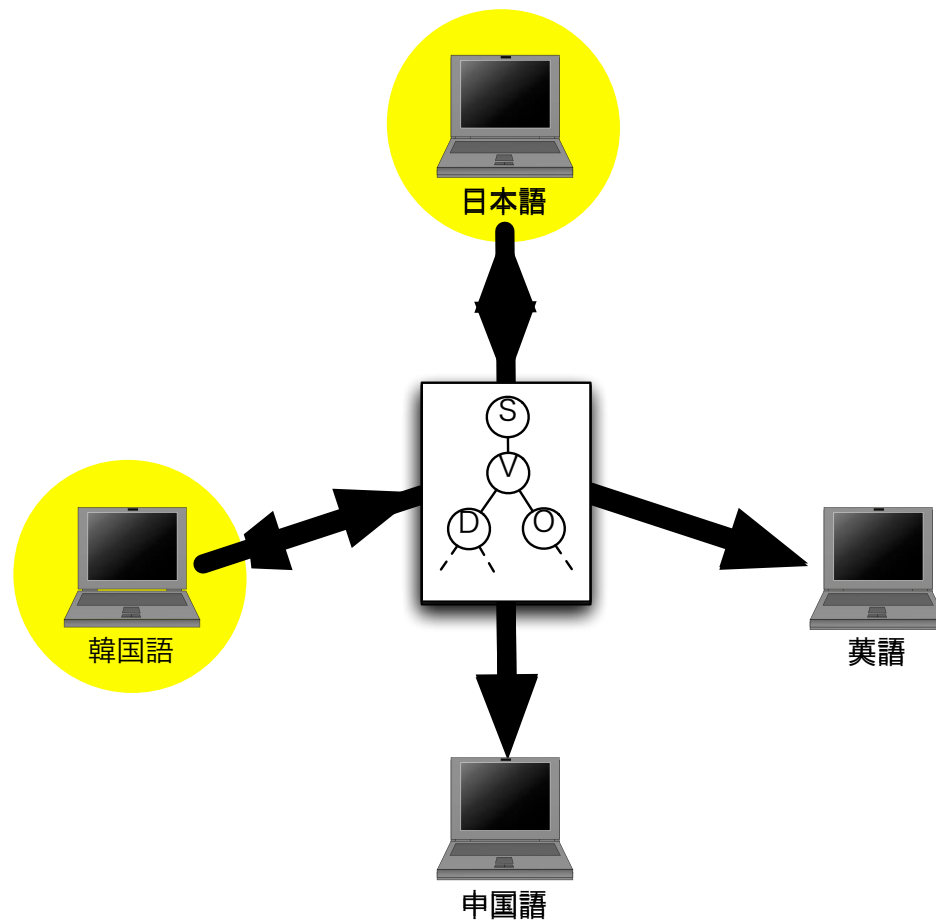
万人の役に立つ機械翻訳

文脈，分野を制限して

自然かつ**正しい**翻訳をする

多言語を一度に扱う

日本語，英語，韓国語，中国語，仏語



ご静聴ありがとうございました

<http://www.musha.com/ipa/>