

産業情報としての特許情報と 産業日本語

2010年2月24日

横井俊夫

Japio特許情報研究所顧問

東京工科大学名誉教授

日本語を鍛える

見直しを迫られる日本

- すがりつきたい日本語

内向きになる日本人→美しい日本語と漢字検定ブーム

(文藝春秋SPECIAL:素晴らしい日本語の世界、季刊秋号、2008年)

- 鍛えるべき日本語

日本語には守るだけの価値がある

→高質国民文学

(水村水苗:日本語が亡びるときー英語の世紀の中でー、筑紫書房、2008年)

→高度経済成長

閉じて守るのではなく開いて守る

→ガラパゴス日本・ガラパゴス化日本語からの脱却

分野に閉じた日本語から分野の外へと開かれた日本語へ
分野内情報と分野外情報との連携効果を高める

日本に閉じた日本語から諸外国へと開かれた日本語へ
日本情報と世界情報との連携効果を高める

人に閉じた日本語からICTに開かれた日本語へ
人の情報処理とコンピュータの情報処理との連携効果を高める



専門家の役割:

閉じた情報の占有者、ジャーゴンの通訳者ではなく
→高価値情報を制作し、活用を先導する者へ

閉じて守るべき情報:

守るべき分野固有情報・守るべき日本情報・守るべき人間情報
→日本としてのインターネット戦略

(岸博幸: ネット帝国主義と日本の敗北ー搾取されるカネと文化、幻
冬舎新書、2010年)

日本語と日本情報

社会・個人の情報環境基盤を構成する言語

日本語⇒日本情報

(日本人が創り・蓄え・伝え・使う情報)

医療日本語⇒医療情報

司法日本語⇒司法情報

産業日本語⇒産業情報

学術日本語⇒学術情報

科学技術系学術日本語⇒科学技術系学術情報

人文社会系学術日本語⇒人文社会系学術情報

外国人日本語教育日本語⇒外国人のための日本情報

日本人の思考言語としての日本語⇒日本人外国語教育

日本産業の情報環境基盤を構成する産業日本語

産業日本語⇒産業情報

⇒金融・経済情報(決算・財務報告書等)

⇒商品・販売情報(利用者マニュアル等)

⇒製品・製造情報(設計・仕様書等)

⇒研究・開発情報(研究論文・資料等)

⇒資源・環境情報(調査資料・データ等)

⇒人材・教育情報(テキスト等)

⇒知財・法令情報(特許文書等)

産業日本語に至るまでに

言語処理と制限日本語

日本語の特殊性→高度な日本語(自然言語)処理技術

吉田 将、田中武美、松山晶子:科学技術文書のための日本語
文法の制限について、情報処理学会自然言語処理研究会
(1983年)

長尾 眞、田中伸佳、辻井潤一:制限文法にもとづく文章作成支
援システム、情報処理学会自然言語処理研究会(1984年)

↓

JUMAN(松本裕治・奈良先端大)

KNP(黒橋禎夫・京大)

等々

日本人国語教育と外国人日本語教育

国語教育

日本語を知っている人への日本語運用能力強化→語彙と情感表現

論理的文章力教育と日本語の仕組教育の重要性

日本語力の低下とテレビ、ゲーム、携帯電話・メールの影響

(毎日新聞:山形大学「文章のいろは」必須科目に、平成22年2月8日)

外国人日本語教育

欧米言語に対して遅れた取組み

日本語を知らない人に日本語の仕組を教える

外国人向け日本語の仕組教育の新しいシステム

日本語能力試験(日本国際教育支援協会)

1984年:15カ国・地域で7千人が受験

2006年:47カ国・地域で43万人が受験

(日本語学習者:国内15万人、海外300万人)



言語学の進展

現代日本語書き言葉均衡コーパス(国研)

日本語教育基盤情報センター(国研)

言語政策とテクニカルライティング

「地球時代の日本語を考える」研究委員会((財)地球産業文化研究所、1991年～):木下是雄(座長)、大宅映子、小塩節、城阪俊吉、小林和夫、堺屋太一、清水洋一、鈴木孝夫、西尾圭子、八戸信昭、丸山理、水谷修、諸見昭、横井俊夫

(木下是雄:言語政策の確立を、中公文庫「日本語の思考法」、pp.68~99(2009年))



日本のマニュアル

日本のマニュアル大賞(毎日新聞社、1992年～1998年)

日本マニュアルコンテスト(テクニカルコミュニケーター協会、1999年～)

日本語テクニカルライティング

テクニカルコミュニケーション技術検定試験(テクニカルコミュニケーター協会)

産業日本語とは

産業日本語→言語学、言語処理、産業情報の融合

産業日本語→人にもコンピュータにも明晰な日本語

人に明晰であるとは→容易に書き下し、容易に読み取れる

コンピュータに明晰であるとは→間違いなく高度に処理できる

↓

2つの明晰性にどう折り合いをつけるのか

かつての制限言語(特に、制限英語)では

能力の低い人に理解できるように制限→コンピュータにはいたって
非明晰

能力の低いコンピュータに処理できるように制限→人にはいたって
非明晰

これからの産業日本語では

能力の高い人(その分野の専門家)にとっての軽微な制限

構文に関する適切な制限、語彙は原則無制限→コンピュータにも明晰

能力の高いコンピュータが対応するため制限は軽微

インタラクティブな処理、協調知の利用→人にも明晰

能力の高いコンピュータ

→高い日本語処理能力、しかし、完全ではない

形態素解析:99%、固有名解析:88%

構文解析:90%、格解析:85%

照応・省略解析:40~60%、談話構造解析70~80%

(黒橋禎夫・ALAGINセミナー)

↓

ブラックボックス型ではなく、インタラクティブ型に

処理結果を明晰に提示→人が容易に適不適を判断

不適結果に別の選択肢→人が容易に変更を指示

↓

効果的なインタラクション

→高いインテリジェンスという裏付け

インタラクション(人との協調対話)には、高い日本語処理能力

人に明晰→インタラクションが明晰

人の想定内に収まり、高い正答率

大量の事例データからの自動学習

協調知の環境

ユーザフレンドリーなインタフェース

コンピュータに明晰→インタラクションが簡明

コンピュータの想定内に収まる妥当な処理結果

柔軟なパラフレーズ機能

明晰な内部表現形式

協調知や自動学習による漸進的機能向上

産業日本語→ひとつではなく目的ごとに多数

産業情報としての特許情報

特許情報

日本の特許情報 ← 日本語

申請される特許文書(毎年約40万件)

IPDLでの蓄積文献数(総数約7100万件)

審査文書(拒絶理由通知書等)

特許分類体系

関連情報としての学術情報(科学技術論文・資料)

米国の特許情報 ← 英語

欧州の特許情報※ ← 英語、独語、仏語 ※欧州特許庁での公用手続言語

中国の特許情報 ← 中国語

韓国の特許情報 ← 韓国語

.....

国際特許分類(IPC) ← 英語、仏語(正本) 中国語、チェコ語、独語、ハンガリー語、日本語、韓国語、ポーランド語、露語、スペイン語(主な公定訳)

特許分類体系:

実運用されている最大規模の産業オントロジー

IPC(International Patent Classification): 世界共通の木構造分類体系

階層構成: 生物分類(界-門-綱-目-科-属-種)とのアナロジー
分類項目数は約7万項目

FI(File Index): 国内運用のためのIPC拡張体系

展開記号(最下位ノードの詳細化)とファセット分類記号(中間ノードの別分類)

分類項目数は約19万項目

Fターム(File Forming Term): 多観点の分類体系

テーマと観点

約2,600のテーマ、その内の約1,800テーマに観点を組み込んだFタームリスト

協調知コモンズとしての特許情報

情報の公開と共有:

権利占有(発明者の権利保護)の代償として情報公開を義務化
新たな発明と新たなビジネスの誕生を促進

情報の品質管理:

特許庁と関連機関による一括管理
審査制度と資格制度

産業情報・学術情報との連携:

知財本部から知識本部へ
総合学術オントロジーと総合産業オントロジー

グローバルコモンズ:

産業活動のグローバル化とアジア経済の台頭
日米欧特許庁三極体制と中国・韓国特許庁の台頭



課題: 日本のイニシアティブの維持、レガシーとの協調、
コモンズの悲劇(パテント・トロール)

東アジア共同体

共同体言語系と共同体情報



欧州共同体言語系

欧州の言語事情

公用語の多くは印欧語族のゲルマン、ロマンス、スラブの3語派、ラテン文字で表記

多言語主義政策→23の公用語

欧州英語という共同体共通語の議論

母語としての使用比率は独語が18%で最多

話することができる比率では英語が51%で最多

母語に加え2つの言語を話せる言語教育

内部作業言語は、英語、仏語、独語

欧州委員会に翻訳総局、通訳総局→毎年約11億ユーロ(総予算の1%)

公文書を23の公用語に翻訳

欧州特許庁(EPO)

特許公用語は、英語、独語、仏語、年間申請件数は、日本の約40%



東アジア共同体言語系は？