

第6回産業日本語研究会・シンポジウム

2015年2月24日

辻井重男

暗号研究者から見た言語・論理とデジタルフォレンジック 概要

古典暗号時代に、暗号解読とその誤訳などが第1次・第2次大戦の勃発に及ぼした深刻な影響について触れた後、現代暗号研究者から見た組織通信における言語と論理の役割を多角的に考察する。

特に、現代暗号の安全性証明における論理的証明、クラウド環境下における秘匿検索・暗号化状態処理のための自然言語処理と論理学の活用、デジタルフォレンジックや公共情報コモンズ（災害情報伝達システム、Lアラート）における言語翻訳課題の緊急性について考察する。

2014年7月辻井企画講演会の動機

法令工学 北陸先端科学技術大学院大学(JAIST) 21世紀COE

- 論理学と日本語の両面から
- 法律と論理学・言語処理 学際的総合研究
- デジタルフォレンジック 武田薬品 80億円 殆ど翻訳料
- 公共情報コモンズ(マルチメディア振興センター)
- 避難情報のリアルタイム多言語化
- イギリス 17世紀の英語改革 外山滋比古
- 個人的研究動機
- 論理学暗号方式の創始
- 放送通信の4類型と組織通信—情報セキュリティ概念の高度化

日本語の情報力強化 -概念図-

利用分野

産業・ビジネス・特許の国際競争力強化

法令工学（法令文・法令相互間の矛盾検出技術を言語の構造化と論理学の高度化・多様化の両面から高める

デジタルフォレンジック（証拠保存・分析・開示システムにおける高度なデータ検索・探索と正確な英訳）

クラウド環境下における秘匿検索・論理学暗号・秘匿

災害情報システム（公共情報コモンズ・アラート）における多言語翻訳

観光・オリンピックにおける多言語翻訳



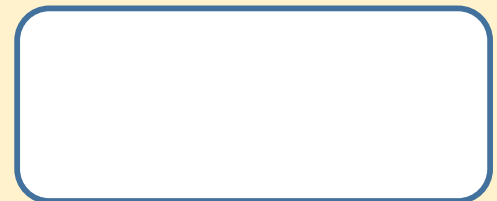
基礎的手法

言語処理・
翻訳技術

論理学の高度化・多
様化

秘匿検索・
秘匿探索・暗号化処
理技術

証明・認証・
証拠保全技術



誤訳

御前会議

Meeting in the Morning

天聴ニ達セラレアリ(難解な表現、文語調)

甚々恐懼ニ堪エサルモ(天皇に対し出すぎたようで恐縮である)

誤訳



重大な懸念を抱いている(政治家自身が)

最新＝最後 → 最終通告

(電文簡単化)

日米戦争回避条件の一つ

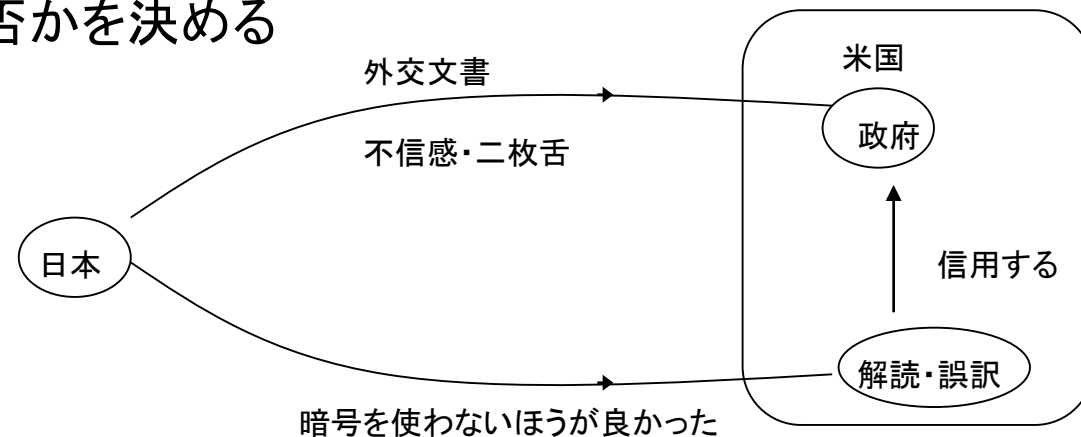
米独開戦時 日独伊三国同盟



日本 自主的に参戦か否かを決める



Automatically と誤訳

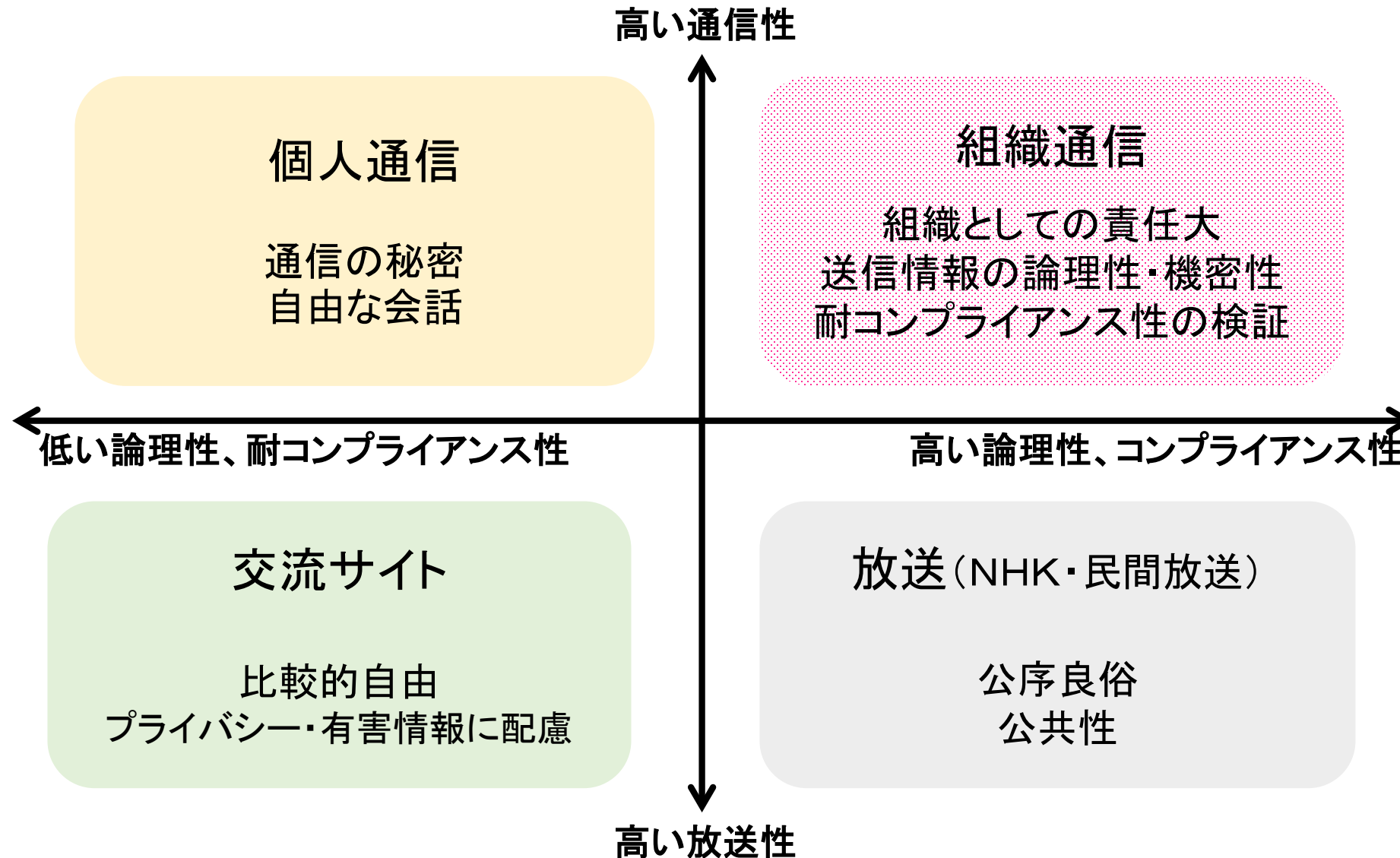


議論の流れ

- 1. 情報伝送・暗号理論研究の立場から
- 2. デジタルフォレンジクと日本語の課題
- 3. 日本語の論理性に対する私見
- 4. 言語と論理に関する課題
- 付録 1. 放送・通信の4類型と情報セキュリティ概念の高度化―第2報
―組織通信と公共情報コモンズ(Lアラート)
- 2. 尸位素餐（しいそさん）とは？

1. 情報伝送・暗号理論研究の立場から

放送・通信の4類型と組織通信



組織間通信における機密通信の必要性

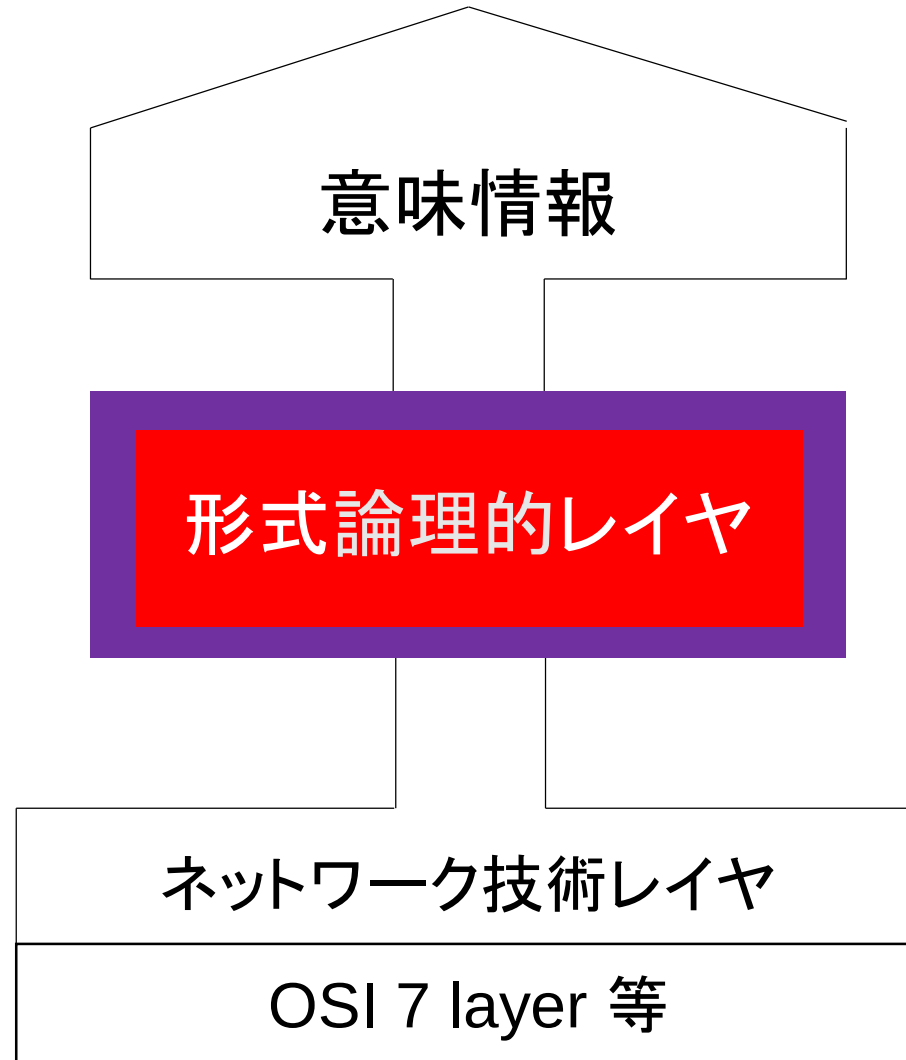
- 情報量の爆発的増大、組織の複雑化、
- 関与者 (Multi-Stake Holder) の拡大を背景に
- 企業、自治体、医療・介護ネットワークなどの
- 組織間における高度な機密通信の必要性増大

例：My Number、税と社会保障、収入・資産

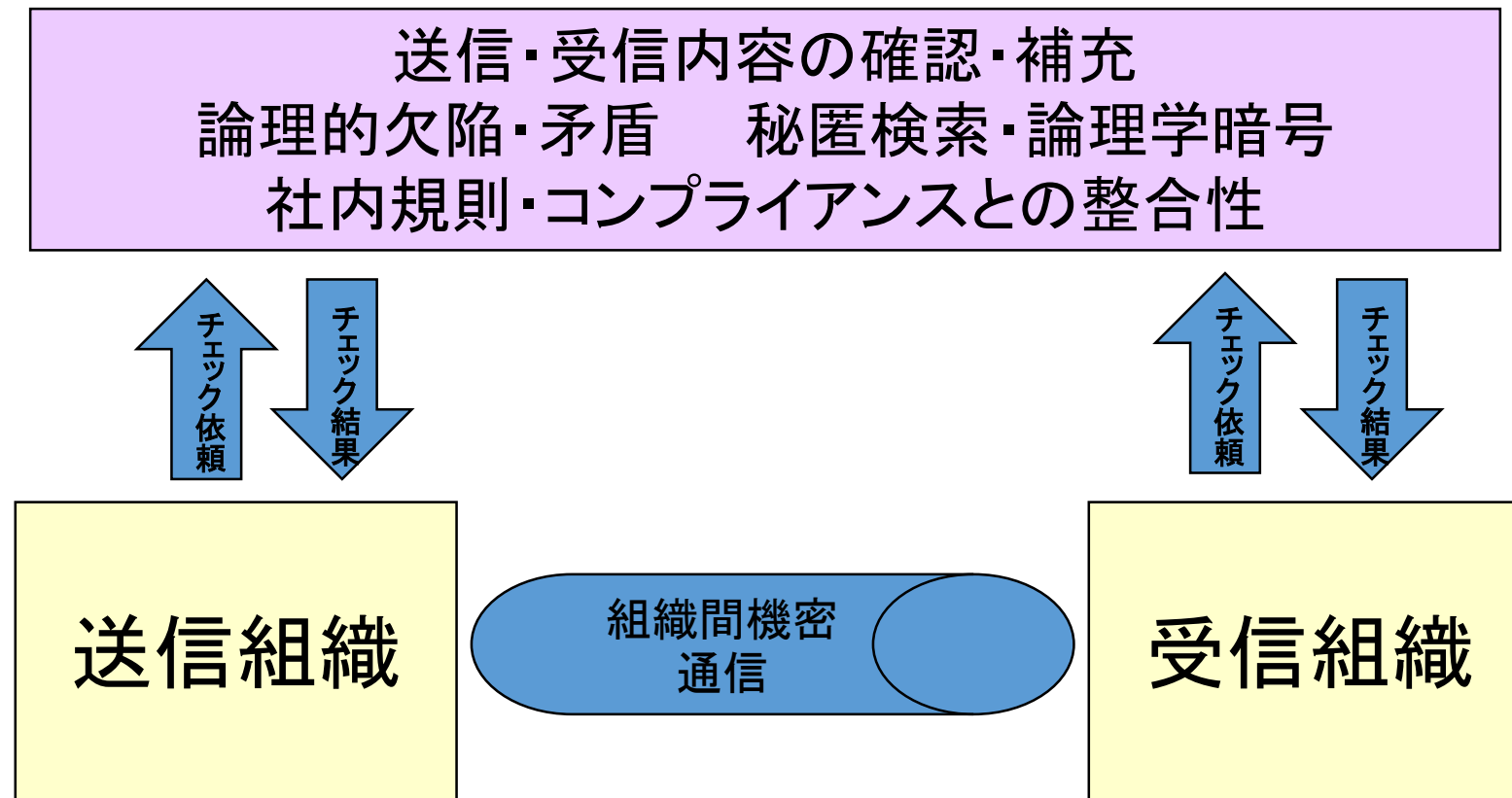
全自治体(居住区域外)に跨る個人情報把握の必要性。

プライバシー保護のため、必要な個人情報のみ把握。

情報通信の新階層



Open data, Cloud, Big data, My number環境下での 情報通信・セキュリティ概念の高度化への対応



組織間機密通信

- 従来は自治体や病院などの単独組織内で、情報を共有・処理することが多かった。



- ネットワーク化、OCBM環境化、関与者の増大

例えば、医療・介護組織間の連携・業務指示

- 事業所単位での活動から連携へ
- 作業指示書などが書類でなく情報通信で
- ① 機密・プライバシーを保護し、
- ② 情報の論理性・耐コンプライアンス性を高めつつ、
複数組織間や民間・行政間での情報交換の増大

組織通信に要求される機能

正確性、信頼性、緊急性、迅速性、証拠性(デジタルフォレンジック)

機密性(組織暗号、構造化言語による秘匿検索)

情報限定性、論理的無矛盾性、

法的整合性(耐コンプライアンス性)

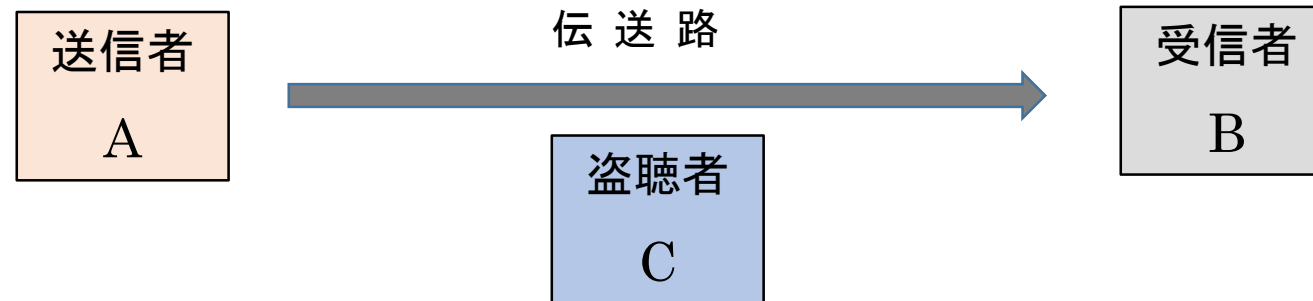
多言語性 (赤字は、本委託研究で実施中)

法令工学

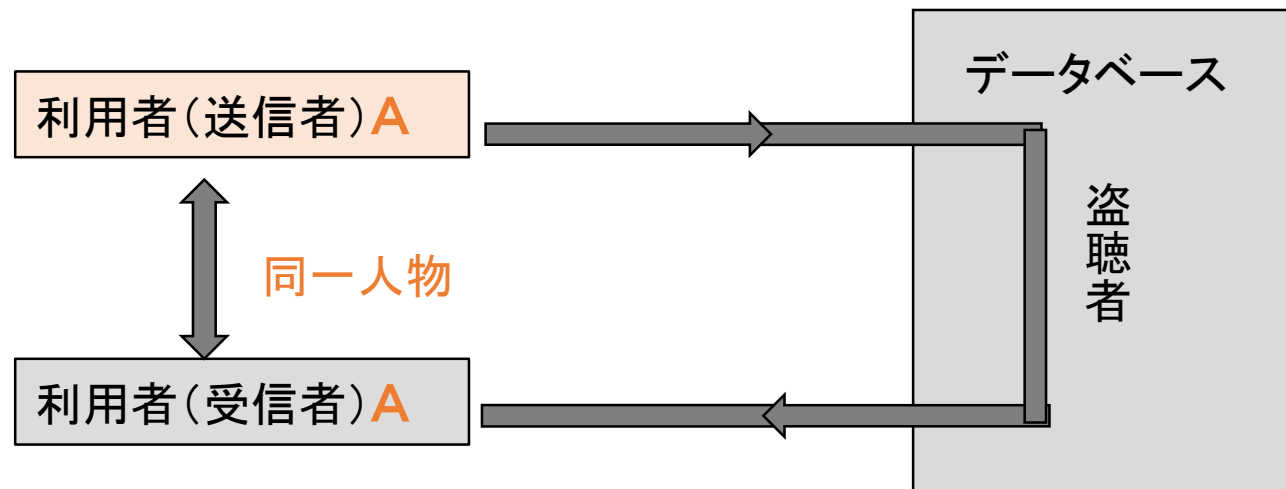
JAIST 21世紀COE 法令工学の提案 片山 卓也

- 「法令工学とは、法令が
 - 1) 制定目的にそって適切に作られ、
 - 2) 論理的矛盾や文書的問題がなく、
 - 3) 関連法令との整合性がとれていることを検査・検証し、
 - 4) 法令の改訂に対しては、矛盾なく変更や追加削除が行われること
- を情報科学的手法によって支援することを目的とする学問分野である」
- 法令工学でいう法令とは、法律のみでなく、都道府県の条令や、企業の社内規則なども対象にしている。
- 論理学と自然言語処理の相性を両側から高める研究が重要

秘密通信と秘匿検索



(a) 秘密通信の場合

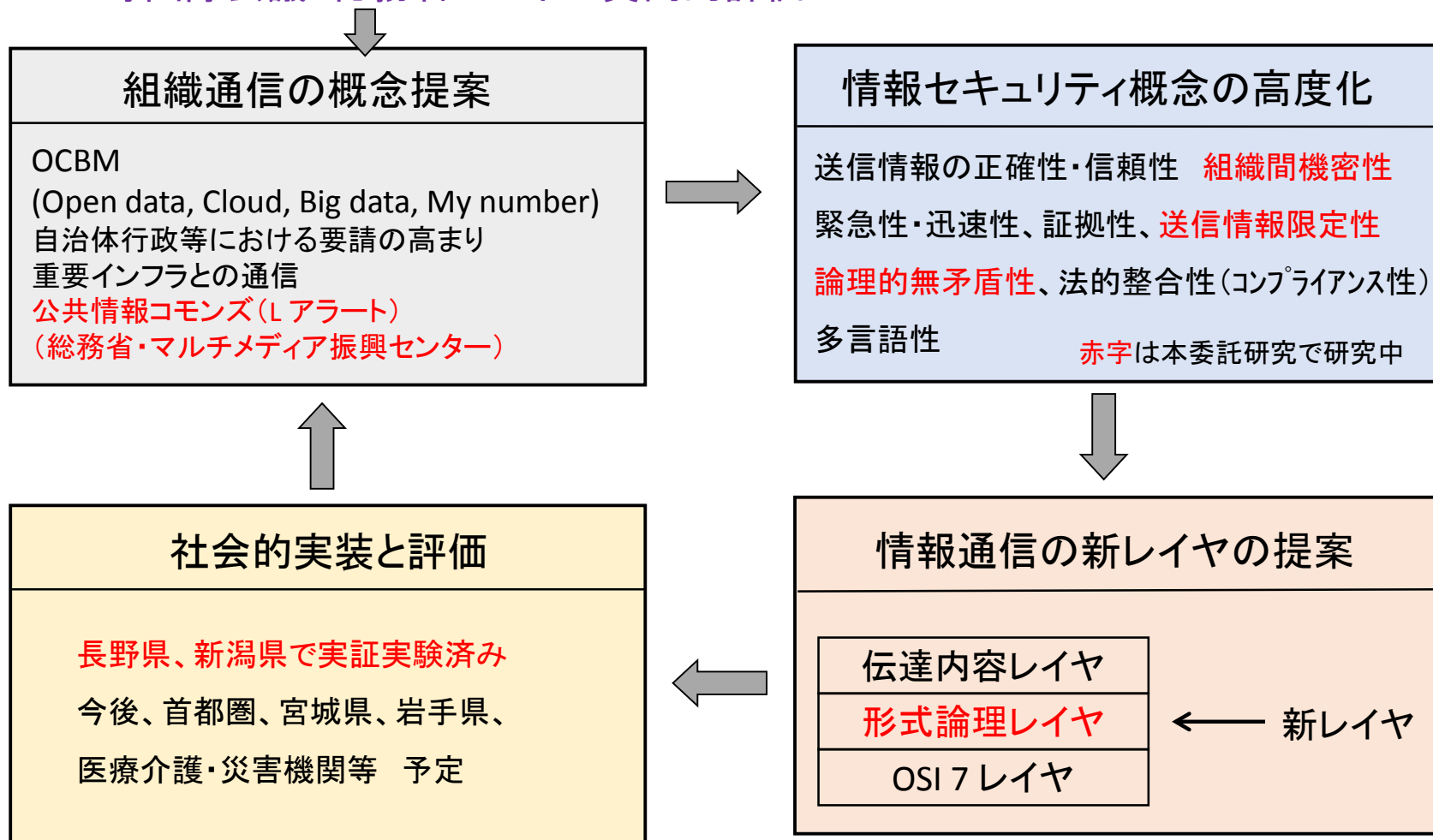


(b) 秘匿検索の場合

自然言語から論理式への変換(暗号化)については伝達には必要はない

組織通信概念の提案とその実現に向けて

IEEE等国際会議・総務省シンポで賛同的評価

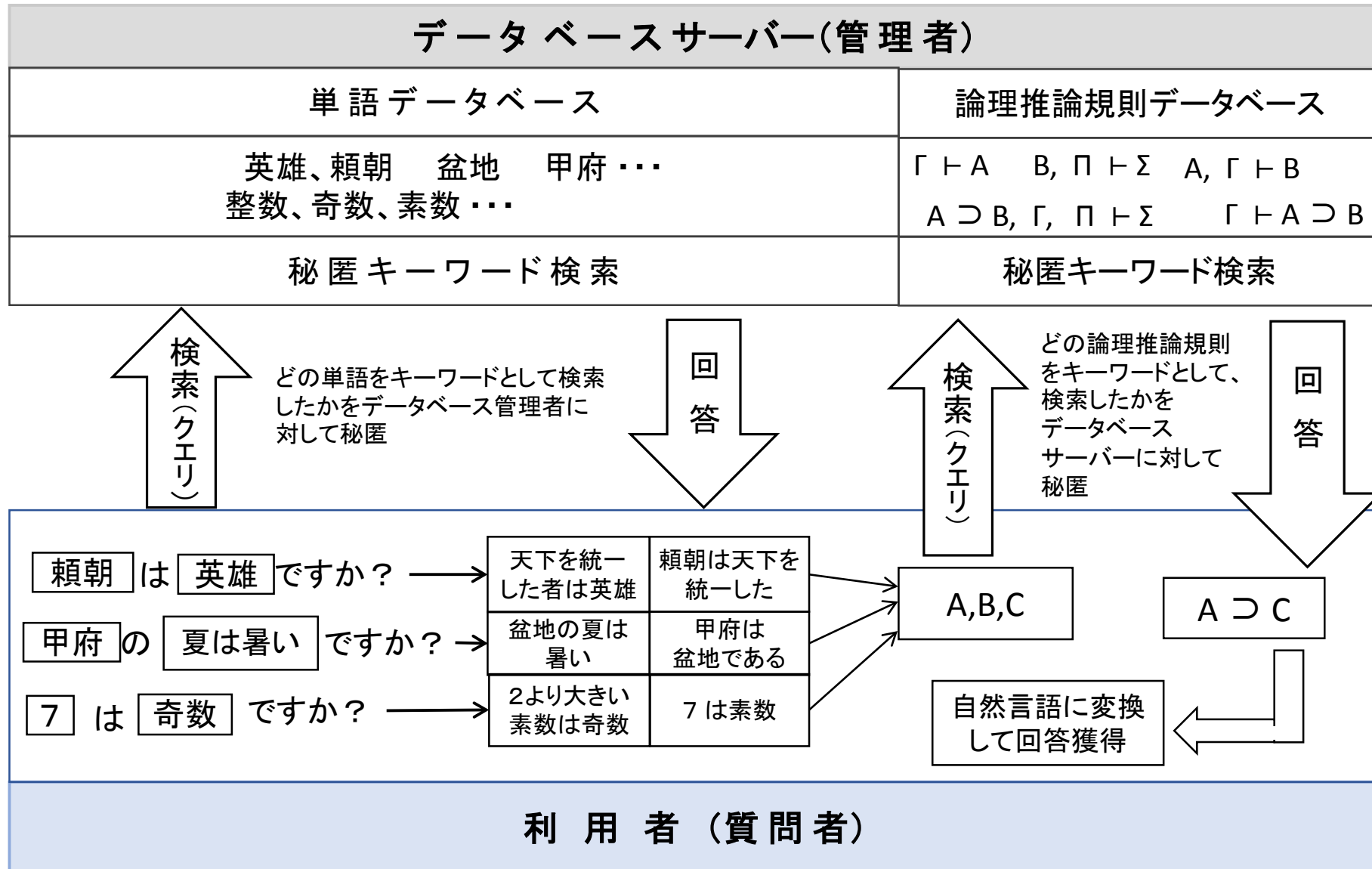


通信の意義

高速大容量→信を通わす

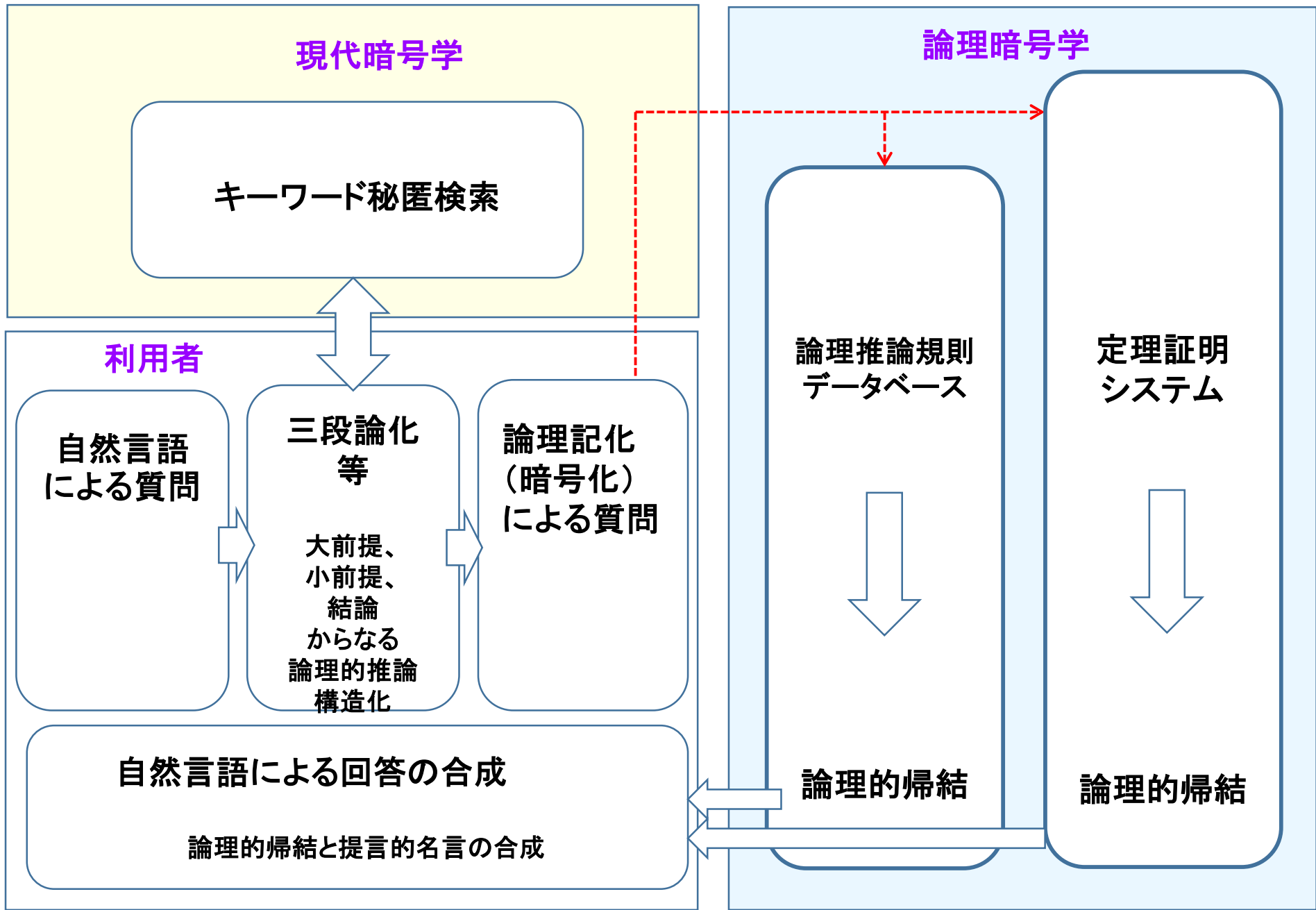
- OCBM時代 Open,Cloud,Big data,My number
- ビジネス、法令、科学技術文書などの論理性
- 送信内容の論理的欠陥・矛盾の検出 CIA概念の拡大
- 質問・回答 証拠の論理的保全・開示
- 暗号化 平文を論理式に直接暗号化
- 現代論理学を利用 古典論理を超えて
- 6月2日 日経朝刊 米国での訴訟和解 深刻な課題
- デジタル・フォレンジック 証拠の保全・開示
- 武田薬品工業 80億 殆どが 翻訳料

現代論理学暗号に基づく 秘匿検索・回答文作成システム



アリストテレスの三段論法のモデルの例

- 言い換えれば, モデルが開示されなければ, 何を推論しているのかわからない.
日常言語と導出規則2 ($A \supset B, B \supset C \vdash A \supset C$) は, 多から1への写像である.
- 多から1への写像を説明する「導出規則2 $A \supset B, B \supset C \vdash A \supset C$ に対応する日常言語を使った別のモデルによる三段論法」には次のようなものが考えられる.
 - 甲府は(A), 盆地である(B)
 - 盆地は(B), 夏暑い(C)
 - ゆえに, 甲府は(A), 夏暑い(C)
- 2 より大きい素数は(A),
2 より大きい奇数である(B)
2より大きい奇数は(B), 7である(C)
ゆえに, 2 より大きい素数は(A), 7 である(C)
- このように, 導出規則2 $A \supset B, B \supset C \vdash A \supset C$ に意味を与えるモデルと導出規則自体を分離して考えることが, 情報セキュリティにおける「他者からの隠ぺい概念」に結びつく.

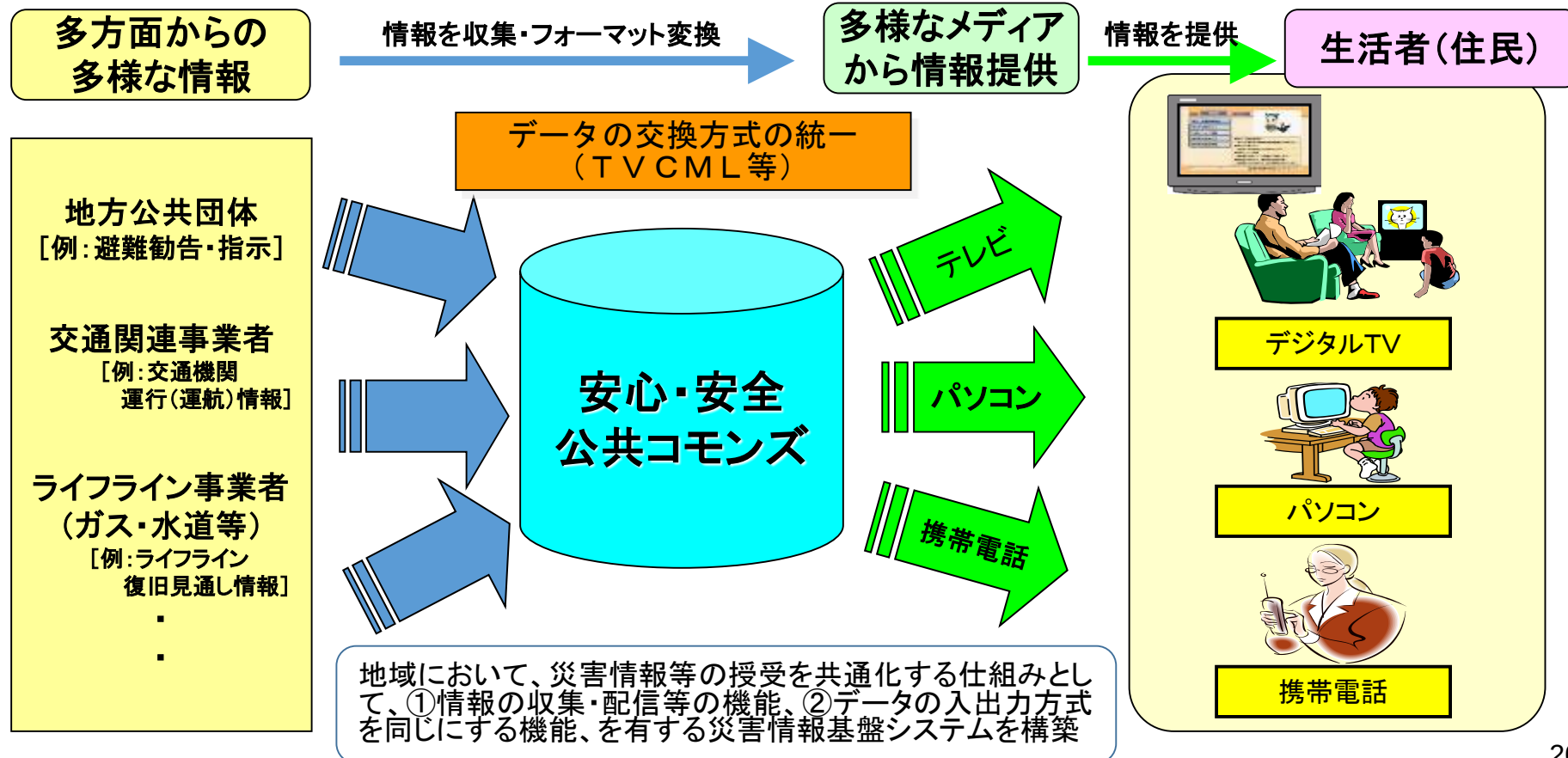


現代暗号と論理学暗号の融合による秘匿問合せ/回答システム

『安心・安全公共コモンズ』のコンセプト (FMMC吉田部長)

多様なメディアを活用して、高齢者をはじめ誰もが、いつでもどこでも、分かりやすい形式で地域の安心・安全に係るきめの細かい情報を迅速に入手できるようにするための具体的な仕組みを、早急に構築することが喫緊の課題

住民への情報提供が容易な『安心・安全公共コモンズ』の構築が必要



2. デジタルフォレンジック ー国際訴訟と翻訳の課題

参考

- 1) 守本正宏著「カルテル・PL訴訟・特許訴訟」ディスカバリ
• (企業家大学出版2012年3月)
- 2) 改訂版 デジタルフォノンレジック事典
• デジタルフォノンレジック研究会編(日科技連)
• 辻井; 日本語の論理性向上のための論理学・自然言語処理
• ・機械翻訳(第10章 10. 1)

米国における 民事訴訟—Pretrial

- 6月2日 日経朝刊 米国での訴訟和解 深刻な課題
- デジタル・フォレンジック 証拠の保全・開示
- 武田薬品工業 80億 殆どが 翻訳料
- -----
- 日本企業 不必要な資料まで翻訳 情報漏洩

「ディスカバリ」 守本 正宏著

10周年記念 表彰普及・啓発賞

- 日米民事訴訟比較 米国文化 当事者主義；
- プリトリアルの段階で、被告と原告双方が、徹底的に「証拠開示（ディスカバリ）」する。
- デジタル署名・公開鍵暗号の役割
- 余談；池上彰番組 エニグマ機 出演 しかし・・・
- 相手側が提供した情報の中から証拠を見つけ出す。
- 日本本社の在米国子会社が訴訟にあった場合の
- 翻訳の費用は莫大。
- 機械翻訳・自然言語処理、現代論理学の導入

e ディスカバリの流れ

- 1) 特定 (identification) 関連するデータ・保有者などのデータマップ
作業範囲・証拠保全計画策定
- 2) データの保全 (Preservation) litigation hold
- 3) データの収集 (Collection) 複製データの作成
- 4) データプロセス (Processing) 無関係な情報を除外
課題; 日本語処理 必要・不必要?
- 5) 分析 (Analysis) 高度なキーワード検索・PPDM・翻訳の絞込み
- 6) 証拠閲覧 (Review) 人手によるデータ選別 タグ付け (coding)
- 7) 提出ドキュメントの作成 (Production) 提出用ファイル形式に変更

1) 特定 (Identification)

- e-Discovery作業の対象となる機器とデータを具体的に確認、整理、確定する作業。
- データの所有者に対しては、業務内容、及び、データの保管状況の確認などを中心に、ヒアリングを行う。
- 現状確認・情報整理終了後、作業範囲 (Scope of Work) を定め、証拠保全計画を策定

Custodian と保管場所 (パソコン、サーバー、USB, CD-R, eメール等) の確認

2) データの保全 (preservation)

- Litigation Hold ;
 - 訴訟に関する全てのデータの削除・変更・改竄を防止
 - 複製元と複製先のハッシュ値比較
 - 違反の場合、 高額賠償など
- 3) データの収集 (collection)

4) データプロセス(Processing)

弁護士の指揮下で提出すべきデータか否かを選別

- 不利な証拠も提出義務
- 有利な証拠も漏れなく

• 5) 分析(Analysis)

- **注意** ディスカバリ業者の選定を誤ると
- 日本独自のアプリ(メールソフトや圧縮ファイル等)
- 絞込み精度低下→ 翻訳費用膨大

6) 証拠閲覧 (Review)

人手によるデータ選別

視認によるタグ付け (Coding) 最も時間と費用を要する。

- (1) 訴訟に無関係
- (2) 関係はあるが、秘匿特権が認められるもの
- (3) 訴訟に関係があり、提出すべきもの
- (4) 提出すべき重要

数十人のレビュアーが何ヶ月ということも

7) 提出ドキュメントの作成 (Production)

私案

- 日本語への翻訳を極小化することが不可欠
- そのため、
 - (1) 分析過程(5.)で、より高度な論理学を導入してはどうか。
 - (2) 法令工学との連携を図ってはどうか。
(法令データベースを参照・知見を得る)

3. 日本語の論理性に対する私見

- 尸位素餐（しいそさん）とは？ 国家の命運
- —ビッグデータ時代の日本語の論理性と国際性
- デジタルフォレンジックでも誤訳の心配 企業生命
- 付録参照

日本語と日本人のDNA

- 長いトンネルを抜けると雪国であった 川端康成
- The train came out of the long tunnel into the snow country

サイデンスティッカー

その列車が、その長いトンネルを抜けると、その雪国に入った
再日本語訳

主体一客体？ 純粹経験 主語・定冠詞 不要？
自他一体の境地

自我の形成と言葉

フランスの2歳の子供でも (写真)

西田幾多郎著 「日本文化の問題」(1940年、昭和15年)

- 「私は日本文化の特色と云うのは・・・どこまでも自己自身を否定して
- 物となる、物となって見、物となつて行ふと云うにあるのではないかと思う。己を空うして物を見る、自己が物の中に没する、・・・
- 日本精神の真髓は、物に於て、事に於て一になると云うことでなければならない。」
- 自己を否定し、私を排し、無心になって、対象と自己を一体化する。
- (佐伯啓思著 西田幾多郎 無私の思想と日本人) VS 加藤尚武

日本語、途方もなく自由だった？

日本語が持つ「途方もない融通無碍な自由さ」だ。「非論理的なものも『てにをは』がつなげてしまうなど意味を超えて感情を喚起する、ある種の分泌性がある」。そして日本語を操る我々にも「つじつまが合わないものを受け入れ、そこに美や叙情を感じる性質がある」というのだ。

（赤田 泰和）

朝日新聞 2013年4月30日 夕刊掲載記事より
結構ですね。

日本語の課題

- 揺れ動く感情を連綿と綴る
湿度100%の演歌？（広田）
- 外山 滋比古
- 150年振りに改革必要
- イギリス 17世紀 大改革
- 情緒・感覚→明晰な論理表現へ

イギリス・ロイヤルアカデミーの英語改革

(山口 浩 作成)

『今の英語』

- ◎詩、文学など
情緒・感情を中心とした
曖昧模糊たる表現が多い
- ◎これではイギリス人の新しい
考え方を十分に表現できない

英語改革

『改革した英語』

- ◎科学文章など、科学的、論理的、
思考的な表現に重点を置く
- ◎これなら、イギリス人の納得いく
内容を明晰に表現できる



イギリス・ロイヤルアカデミー

トーマス・スプラット
(1635-1713)

The History of the Royal-Society of London
for the Improvement of Natural Knowledge



4. 言語と論理に関する課題

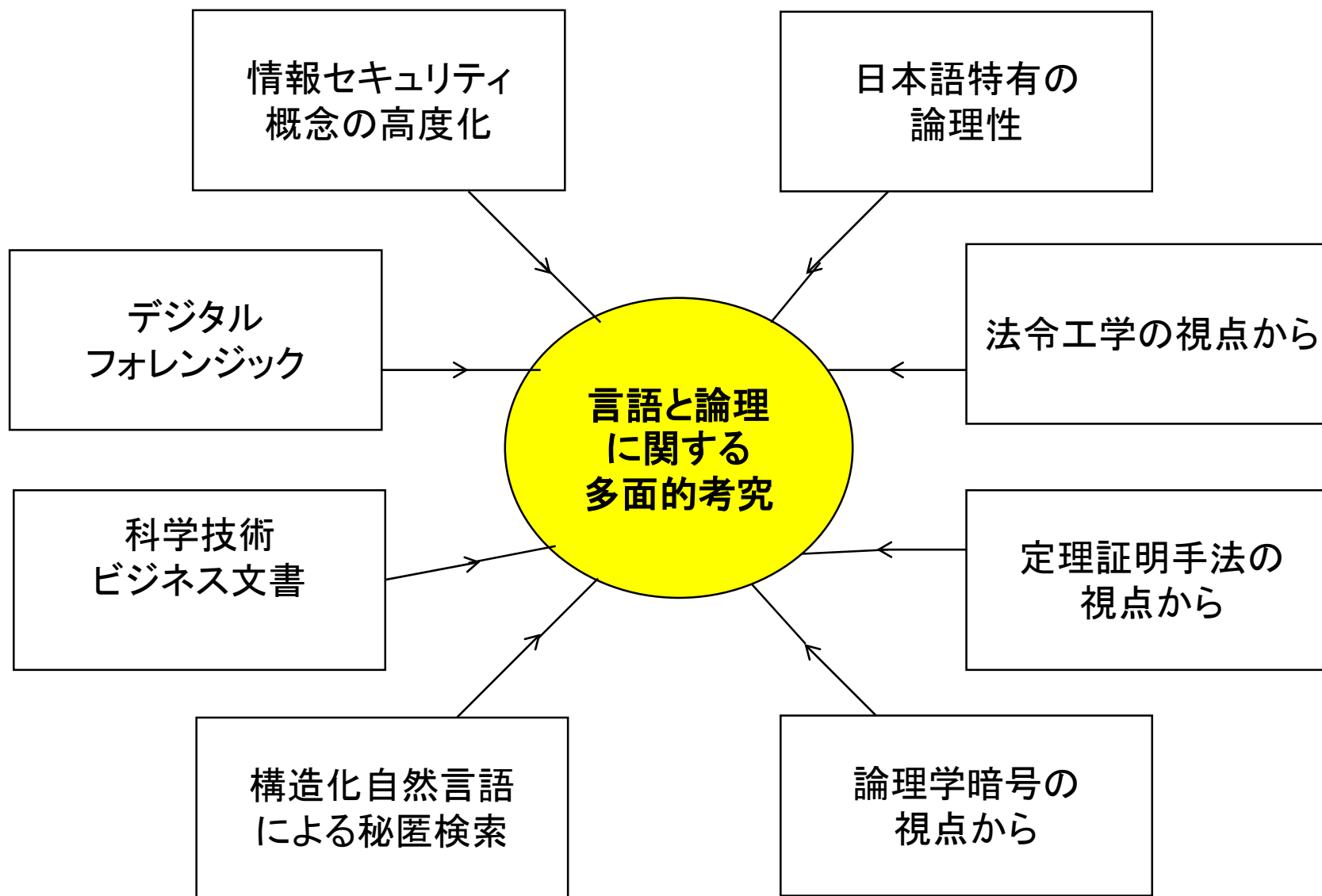


図 言語と論理に関する多面的考究の必要性

自然言語から人工言語へ

与謝野晶子は与謝野鉄幹を愛している。

与謝野鉄幹は与謝野晶子に愛されている。 OK

皆は誰かを愛している。→ 誰かは皆に愛されている。 OK？

「いかなる行為にも、その究極的目的がある」 →

「ひとつの究極的目的を目指して、全ての行為がなされる」 OK？

アリストテレス ニコマコス倫理学

「感覚的な事物の存在は、全て何らかの精神に依存している」 →

「ある精神が、全ての感覚的な事物の存在に不可欠である」 バークレー OK？

(飯田 隆;言語哲学大全 I～IV)

自然言語では隠されている構造→論理形式が記号表現の表面に現れている言語

命題論理から述語論理へ

- 命題論理(文を単位とする推論)やアリストテレスの三段論法では、主語・述語の関連が消えてしまって、推論の正当化が出来ない。
- アリストテレスの方法
- 「全てのAはBである。」 項に一定の記号を与える。
- 「全ての」とか「ある」の部分はそのまま温存。
- 正当化には 文を構成する要素部分の取り扱いと普遍量化子・存在量化子の導入が不可欠
- → 述語論理

述語論理 項からモノ(変数)へ

- パンダは笹を食べ、コアラはユーカリを食べる。
- パンダもコアラも動物である。
- 笹もユーカリも植物である。
- -----
- 質問; それならパンダもコアラも草食動物ですか。

記号表現

$F(x)$; x はパンダである。 $G(x)$; x はコアラである。
 $H(z)$; z は笹である。 $F_1(x)$; x はユーカリである。
 $G_1(x,z)$; x は z を食べる。

- $\forall x (F(x) \supset \exists z (G_1(x,z) \wedge H(z)))$
- $\wedge \forall y (G(y) \supset \exists z_1 (G_1(y,z_1)$
- $\wedge F_1(z_1)))$

自然言語と論理学の相性は悪いか？

- 富美子は物理の得意な女性だ。
 - 富美子は女性だが、物理が得意だ。
 - 富美子は女性のくせに、物理が得意だ。
-
- 論理学は ニュアンス・感情表現は苦手？
 - 科学技術、法律、ビジネス等に限れば
 - 可なり相性を良く出来る。

今後の展開 多様な論理学の導入が必要

1階述語論理・古典論理から多様な論理へ

神の視点(古典論理); 有罪か無罪かは決まっている?
判決の前に死亡するかも?

人間の視点; 法律や医療の相談では、**直観主義論理**、**様相論理**

など多様な論理学を導入する必要がある。logics

数学においてすら、構成主義、直観主義の立場がある($\pi=3.14\dots\dots$)

暗号の安全性証明 IND-CCA等では専ら古典論理(背理法)だが。

親が叱らない と 子供は勉強しない→子供が勉強する と 親は叱る?

古典論理・命題論理から非古典論理・述語論理へ

- 2重否定＝肯定 を認めない
- × 排中率、背理、対偶、含意、ド・モルガンの一部
-
- Q 華子さん 好きですか？ A 好きでないことはありません
- Q それでは好きなのですね A ？
- 親が叱らない→子供は勉強しない 子供が勉強する→親が叱る？

多値論理、直観主義論理、様相論理・・・

JAIST 21世紀COE 法令工学 Davidsonian style

フレーゲ(1848－1925)の命題観

- 「伊藤 邦武著;物語 哲学の歴史」より
- 命題とは関係を表現する関数が、その定項や変項として個別的な対象を含むものであり、それに加えて、それらの対象についての量的表記を行う「量化子」を加えたものである。
- 例;師弟 という関数が ソクラテス、プラトンという値をとる。
- 新しい論理学は主語－述語の命題形式を廃棄することで、
- 命題同志の演繹的な関係の推論の力を飛躍的に増大。
- → 形而上学の大きな変革へ(意識の哲学から言語哲学へ)

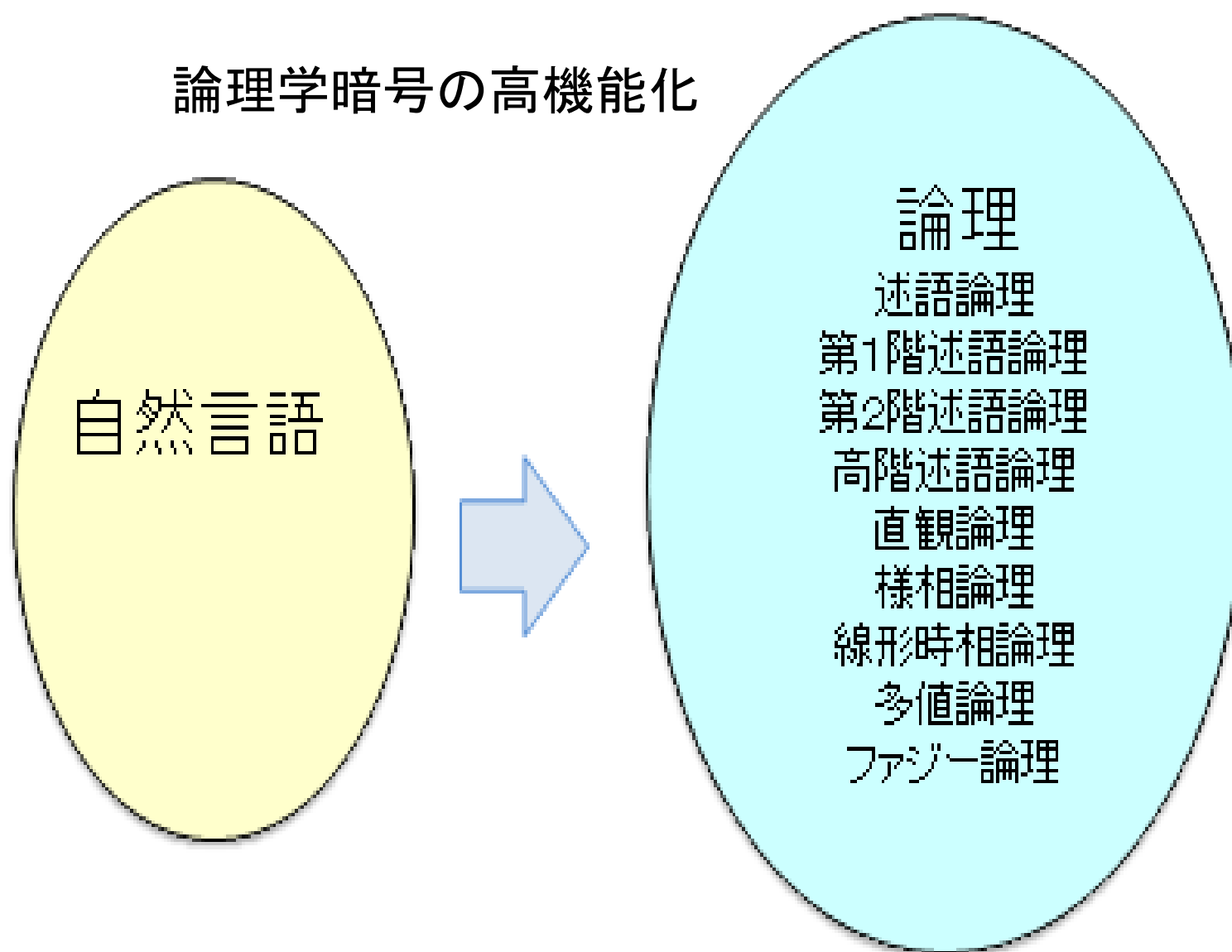
古典論理から非古典論理へ

- 古典論理＝神の論理 nearly ＝数学の世界
- 排中律 2重否定＝肯定
- 真か偽か 善か悪か 好きか嫌いか 有罪か無罪か
- 人間世界 好きでないことではない＝好き？

非古典論理

- 直観主義論理 × 排中律、2重否定＝肯定
- 田中角栄 有罪？
- 様相論理 知識、信念、自覚 例；医師 薬100mg communication
-

論理学暗号の高機能化



長期的重要課題

- ー日本語文の論理性向上のための論理学・自然言語処理・機械翻訳ー
- (辻井; IDF事典原稿ー将来の課題より抜粋)
- (1) 日本語の論理性を高めること
- (2) 現代論理学と日本語を含む
- 自然言語処理の親和性を高めること