3E3-1

知的エージェントとその言語発達に関する研究フレームワーク

A framework of intelligent agents and language development

新田 恒雄
Tsuneo Nitta

豊橋技術科学大学 大学院工学研究科
Graduate School of Engineering, Toyohashi University of Technology

Robots, or intelligent agents, and humans will exchange knowledge and learn each other in near future. Aiming at such a symbiotic community, we have been developing a framework that enables the agents to acquire language and task skill at a real world, to develop knowledge by oneself, and to communicate each other in the community. In this paper, firstly the structures of intelligence both without and with language are discussed, and then after describing the representations of speech and image objects, the acquisition of word meanings and interaction strategies is introduced. Illogical learning biases are also discussed.

1. はじめに

知能ロボット研究のロードマップが示され[1], 多用な分野に亘る課題と解決方策が議論されている。ロボットと人間との対話は, 現状では定型的で表面的なものに留まっている。近未来の知能ロボット(知的エージェント)が, 実世界から自律的に知識を獲得し, 人間と対話できる能力を持つには, ロードマップが将来課題に挙げる, フレーム問題と記号接地問題を視野に入れた研究が不可欠になる。

一方, ネット上に日々膨大なデジタル情報が蓄積される時代を迎え, 個々人が情報を知識として利用するための研究が, 多方面から進められている[2]。しかし情報の知識化は, かつて N. Wiener が指摘したように個人毎に異なる形をとる。一つの観点から貼られたリンクが, 別の人にとっても最適であるとは限らないため, 適切なリンクを探す作業は個人に多くの手間を強いる。Web2.0 の一部技術に観られるように, 全ての人のための一般知識ではなく,「自分にとって意味ある知識」を如何に蓄積し, 利用できるようにするかが重要視されている。筆者は, この解決には個人に適応した知的エージェント(知能ロボット, アバターロボ)の開発が不可欠と考えてきた。

我々は, 知的エージェントの開発を通して, (A) 個人のために情報を知識化し, 対話を通してそれらの利用を支援すると共に, (B) 音声言語を含む相互行為を通して, ロボットと人間が共生する社会を構成することを最終目標とする研究を行っている。このような共生コミュニティを実現するには, まず知的エージェントが言語とタスクスキルの知識を獲得し, 自律的に知識を発達させ, 対話を通して互いの知識を交換することを可能にする研究フレームワークの設計が必要になる。

本報告では, まず 2. で言語知識を持たない場合と持つ場合について, 知の体系を考察する。その中で, 多世界に分散表現された知識と, 状況に応じた知識の利用法を検討する。次に, 3. で音声の表象としての音素弁別特徴, および画像表象について説明した後, 4. では, これまでに我々が行ってきた語意獲得と対話戦略獲得に関する研究を紹介する。また, 最後に, 知的エージェント開発に今後重要と考える四つの研究テーマ, 多世界モデルとその部分空間設計, タスク知識と言語知識の統合, 相手理解, 未知の事物に対する説明能力を挙げた。

連絡先：新田恒雄, 豊橋技術科学大学, 〒441-8580 豊橋市天伯町雲雀ケ丘 1-1,(0532)44-6890, nitta@tutkie.tut.ac.jp

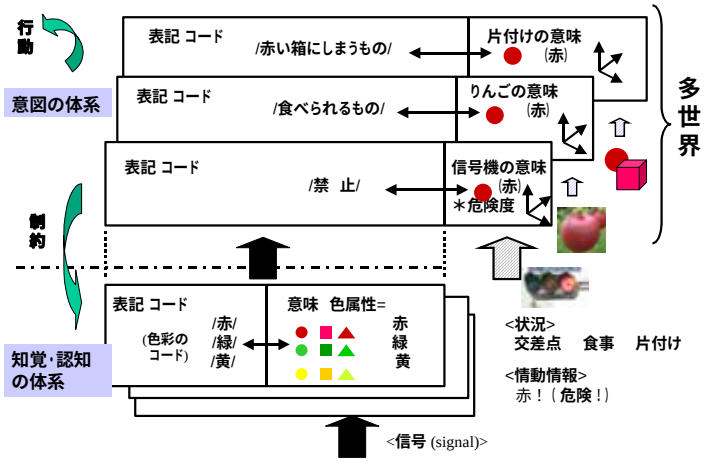

図 1 言語を持たないエージェントの知の体系

2. 知の体系

知的エージェントが備えるべき情報処理機能を考える上で, まず知の体系を検討する。以下, 言語を持たない知の体系, すなわち人間以外の動物も備えるであろう知の体系を考察した後, 言語獲得後の体系を比較検討する。

2.1 言語未獲得時の知の体系

図 1 は, 言語を持たないエージェントの知の体系を示したものである。図の下部は, 実世界を写像した「知覚･認知の体系」で, 右が所謂, 原型の世界, 左が表記コードである。ここの書き方は, (内容は異なるが) バルトがモードの体系[3]で用いた構造化手法を借りている。表記コードとは, 知覚と認知を経て分節化された, (言語以前の) 表象に対するコードを考える。分節化には意味づけされたプロトタイプが必要なので, これは上部 (価値体系を持つ) とのやり取りから, 設計されるとする。コードの例として, 事物が持つ属性 (形, 色, テクスチャ, ･･) から抽出した「赤さ」,「丸さ」,「ざらざら度」, ･･などを挙げておく。

次に図の上部は, 知覚･認知の体系を通して得た情報を, 状況の中で捉えつつ価値を判断したり, 相手がいる場合には意図

を推定する体系である。この後，自身の意図を生成するため，ここを「意図の体系」と呼ぶことにする。必要ならここに自己意識を認めることができるだろう。価値や意図は学習の中で形成されるが，知能ロボットではゴールを達成する，あるいは褒められることを契機に学習が進む。状況では，既得のタスク知識（片付け課題，地図課題，‥）の違いと共に，超分節情報である情動の違いが意図の解釈を変える。「赤い丸」が，或る状況では片付け（タスク）の中でその対象であったり，別の状況では”赤！（ここでは分節情報ではない）”という大声と共に危険を知らせる標識であったりする。意図の体系は，知覚･認知の体系と状況を含めて記述されるとき，connotation のメッセージを伝える。

価値や意図は，状況に依存するため，意図の体系は図に示すように多世界になる[4], [5]。なお，これらの多世界と連携して，下部レイヤの知覚･認知の体系も分化し，制約として働くようになると考えられる。多世界から状況に合致する世界を瞬時に選択する手段が必要になるが，これには図 2 に示すLDA(Latent Semantic Analysis)[6]，もしくはPLDA(Probabilistic LDA)[7]が適用できる。図の縦軸は知覚･認知の体系でコード化された表象（尤度値），横軸は多世界に対応する（図では，言語獲得後の処理を含むよう，単語を要素に入れた）。知的エージェントは，この行列に特異値分解を適用して（直交）部分空間ψ(m)を構成し記憶する（LDAの場合）。これにより，ある状況で刺激Xが与えられると，部分空間との内積演算(X･ψ)から，対応する世界を瞬時に計算できる（ここで述べた計算原理は，基本的にニューラルネットで行われる特徴抽出や分類（あるいは予測）などの演算とも近い）。なお個々の世界内では，その世界kに特有の（限定された数の）表象群から部分空間ϕ^k(n)を構成するため，効率良く知識を蓄えることができる。

	交差点	食 事	片付け	‥‥	地 図	‥
赤 さ	0.9	0.7	0.5	‥‥	0.3	
丸 さ	0.7	0.6	0.6	‥‥	0.3	
‥‥	‥‥	‥‥	‥‥	‥‥	‥‥	
ざらざら度	0.5	0.5	0.6	‥‥	0.4	
‥‥	‥‥	‥‥	‥‥	‥‥	‥‥	
「単語1（赤い）」	(0.8)	(0.6)	(0.6)	‥‥	(0.1)	
「単語2（丸）」	(0.7)	(0.6)	(0.4)	‥‥	(0.2)	
‥‥	‥‥	‥‥	‥‥	‥‥	‥‥	
危険度	0.8	0.2	0.2	‥‥	0.3	
‥‥	‥‥	‥‥	‥‥	‥‥	‥‥	
食べられる	0.0	0.9	0.1	‥‥	0.1	

図 2 多世界から部分空間を抽出する

2.2 言語獲得後の知の体系

図 3 に言語獲得後の知の体系モデルを示した。図の下部は言語を持たない場合と同様，「知覚･認知の体系」に相当する。新たに言語の体系が加わり，言語表記を使用できるようになった。これにより，表記のぶれを正準化し，意味を命題表象（およびスキーマ）として記述することが可能になる。音声は，言語情報を担う分節情報と，意図伝達に有用な超分節情報（韻律）の二つを含む。単語は，画像などのアナログ表象（とその属性）に名前を貼り付け，世界内を記述すると共に，多世界を表現する部分空間に，ラベルという安定した要素を組み入れた寄与が大きい。

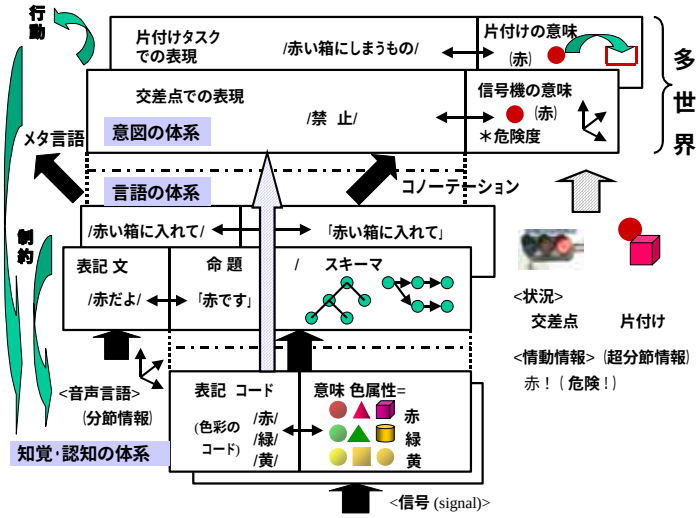

図 3 言語を持つエージェントの知の体系

部分空間から構成される世界では，元の原型を「もの」と考えると，「こと」に相当する。しかし渡辺慧は，この二つがシュミット変換から相互に再現できることをもとに，「もの」と「こと」の相対性を指摘した[8]。

言語の意味を確定するには，命題表象のほか操作などを記述するスキーマを取り入れなければならない。言語を獲得するとエージェントは，

① 相手の行為 （言語表現，非言語表現）と，
② 状況（事態，タスク，‥）を

その都度解釈しながら，意味を理解することになる。しかし，他者との対話理解は，一筋縄では解決しない。筆者は，基本的にディビッドソンの以下の立場をとる[9], [10]。すなわち対話に際しては，話し手の性格･役割などから解釈する“事前理論”と同時に，意味が整合する／しないに係わらず，一定の調整を経て（場合によっては，その場で解釈者に使用させようと意図する，当座理論をアドホックに適用しつつ）合意するとする立場である。このため，エージェントは 4.に触れる対話戦略を獲得しつつ，対話を進行させる能力を持つ必要がある。

なお，言語の体系に知識が蓄積されると，獲得した知識を自分の視点（もしくは社会の視点）で記述しなおす操作が働く。図のメタ言語はそのためのもので，知識を有効利用するのに不可欠なものであるが，ここでは触れない。

3. 表 象

ここでは，知覚･認知の体系における表象について検討する。音声言語の研究分野では，古くから発話とその知覚は一つのシステムなのか，あるいは二つのシステムかが論議されてきた[11]。脳科学分野の最近の研究からは，一つのシステム説に分がありそうである[12]。「聞き手が知覚したいと思っているものは，音そのものではなく，調音の運動である」とすると，言語音声の表象は調音の動作を反映するものが望ましい。Jacobson らは古くに，音素の調音上の差異に音響的･聴覚的差も加味し，全ての言語で使用可能な弁別的特徴(distinctive feature)を提案した[13]。

図 4 に日本語音素に対応する弁別的特徴の一部を示す[14]。また図 5 に発話/aia/ の区間の弁別特徴”high”（舌の位置が高い）の動きを示す。/i/の区間で 1 に近い値を取っていることが見てとれる。図 6 は，全ての日本語音素の弁別的特徴間で，距離を求めた後，多次元尺度構成法(MDS)により三次元布置を描

	a	i	u	e	o	N	w	y	p	t	k	b	d	g
semi-vowel	−	−	−	−	−	−	+	+	−	−	−	−	−	−
fricative	−	−	−	−	−	−	−	−	−	−	−	−	−	−
high	−	+	+	−	−	−	+	+	−	−	+	−	−	+
back	+	−	+	−	+	−	+	−	−	−	+	−	−	+
low	+	−	−	−	−	−	−	−	−	−	−	−	−	−
front	−	−	−	−	−	−	−	−	+	+	−	+	+	−
end	−	−	−	−	−	−	−	−	−	+	−	−	+	−
stop	−	−	−	−	−	−	−	−	+	+	+	+	+	+
voiced	+	+	+	+	+	+	+	+	−	−	−	+	+	+
continuant	+	+	+	+	+	+	+	+	−	−	−	−	−	−
nasal	−	−	−	−	−	+	−	−	−	−	−	−	−	−

図 4 音素弁別特徴 （日本語音素の一部を示した）

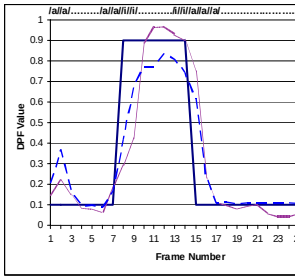

図 5 音素弁別特徴の動作例
発話は/aia/。/i/は舌の
位置が high→ 1 となる。

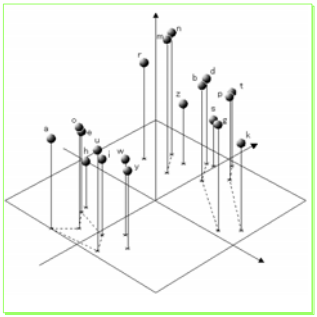

図 6 音素弁別特徴
空間上の音素布置

かせたものである。我々は，弁別的音素特徴(DPF)を抽出する，ニューラルネットの研究を行っており[15]，最新のシステム（RNN+MLN+抑制／強調ネット + HMM(mono-phone model のみ)で構成）では，音素接続や単語接続ルール(n-gram 言語モデル)なしで，90%を越える音素認識率を達成している。DPF は，かなテキストデータと一対一に対応するため，音声合成にとっても相性がよい。

一方，画像に関する表象は，音声のような生成モデルがない。そこで現在は，特徴量として図 7 に示す高次局所自己相関関数[16] を計算して用いている。しかし，この方式は次元数が高いため，これに代わる方法として，画像特徴から弁別的特徴（丸みある／細い／重い／・・・）を定義し，音声と同じくニューラルネットで抽出する方向で検討している。

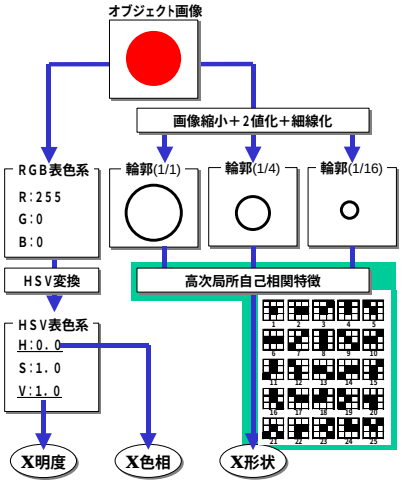

図 7 高次局所自己相関特徴を用いた画像特徴抽出

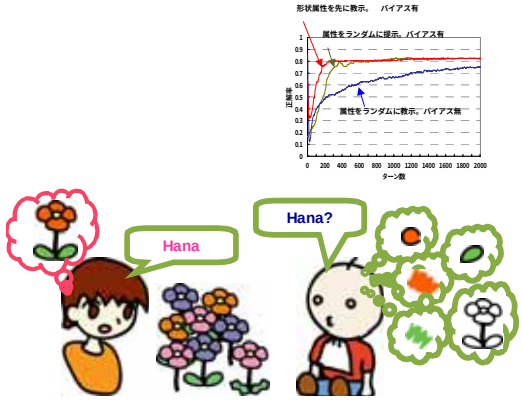

図 8 On-line EM 学習と学習バイアスを
用いた語意獲得[18]

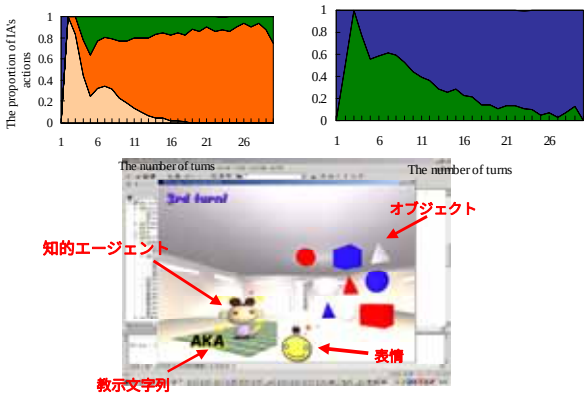

図 9 表情を報酬とした強化学習を用いた
対話戦略獲得[22]

4. 言語獲得 － 実世界に埋込まれた言語の獲得

我々はこれまで，コンピュータ上の画像オブジェクトを対象に，on-line EM アルゴリズム[17]を用いて，語意（形状･色･明度の属性）を獲得する実験を行ってきた。その結果，人間の幼児が持つ様々な学習バイアス（形状類似バイアス，相互排他性バイアス）を組込むことで，学習を効率よく行えることを明らかにした（[18] 図 8 参照）。同時に，実験では言語獲得以前にも，エージェントは世界を観測する中で，属性ごとの基準分布を学習しているという着想をもとに，基準分布と入力との差異から属性を判定し，この判定を利用して単語ラベルとその語意を効率よく学習できることを示した。

また，学習バイアスという非論理的推論の起源について，世代交代を含む簡単な実験を行った結果，対称性バイアス（因果性バイアス）という単一の原理から，幼児の持つ様々な学習バイアスが言語の構造化を通して現れると示唆されることを明らかにした[19],[20],[21]。

次に，対話調整機能を自動獲得させる実験を行った。この中では，獲得した知識を教示し合う際に，質問戦略と教示戦略を強化学習（報酬として表情変化を利用）から獲得させ，語意獲得効率を向上できることを示した（[22] 図 9 参照）。

今後は，以下のテーマを中心に，研究プラットホーム上での実証実験と，ロボットへの実装を進め，関連領域の研究者の

方々と協力しつつ, グラウンディング問題とフレーム問題の緩和を目指したいと考えている。

(1) 多世界モデルと性質の良い部分空間の設計

2.1 に説明した手法に基づき, 多世界を対象とする部分空間 $\psi(m)$, および個別世界（状況）内の部分空間 $\phi^k(n)$を, 発展的に獲得する手法を開発する。

(2) タスク知識と言語知識との統合

知識は, pragmatics→ 意味→ 構文 → 単語という, コンピュータの言語処理手順の逆を辿って獲得されると考える。岩橋が示したグラフィカルモデルベースの手法[23]を踏まえ, さらにタスク知識と言語知識を分離・統合可能な, 協調学習方式を検討したい。

(3) 相手理解

Tomasello は一連の実験結果から, ヒトは他者を自分と同じく「意図を持つ者」として理解し, 他者を通しても新しい語を学習できるとした[24]。この能力は, 語彙爆発のキーと考えられる。実世界から相手の意図体系（世界）を読み取り, 自身の体系から行動を理解するシステムが要請される。

(4) 未知の事物を理解し説明できる能力

状況に応じて, 単語が意味する属性を動的に推論すると共に, 自身が持ち合わせる知識から, 最適な説明を提供できる能力を持たせたい。このテーマは(3)とも関連する。同時に, より効率的な対話を実現する方策が必要である。Vygotsky は, 思考行為としての内言[25], [26]に注目し, その重要性を指摘した。エージェントの発話機能と聴覚フィードバックを利用することを含め（1-model）, 自身の中で知識を自分なりに, しかし積極的に re-write する? 機能を持たせたい。

5. おわりに

知的エージェントが持つべき知の体系を考察し, 知覚・認知の体系で用いられる表象を検討すると共に, 言語獲得の諸段階（語意獲得, スキーマと構文獲得, 対話戦略）に要求される仕様について討議した。

今後は, ここに述べた機能仕様と方式を知的エージェントに組込み, 実証試験を行いたい。

本研究の一部は, 文部科学省科学研究費補助金（基盤研究(C)課題番号 18500130）の援助を受けた。

参考文献

[1] http://www.ai-gakkai.or.jp/jsai/whatsai/rloadmap.html 知能ロボットに関するアカデミック・ロードマップ

[2] 喜連川優, 松岡聡, 松山隆司, 須藤修, 安達淳：情報爆発時代に向けた新しい IT 基盤技術の研究, 人工知能学会誌, Vol.22, No.2, pp.209-214, 2007.

[3] ロラン・バルト：モードの体系,（佐藤信夫訳）みすず書房, 1972.

[4] 三浦俊彦：可能世界の哲学, 日本放送出版協会, 1997.

[5] Kripke, S. A.: Naming and Necessity, G. Harman and D. Davidson (editors), Semantics of Natural Language, D. Reidel Publishing Co., 1972;名指しと必然性,（八木沢敬, 野家啓一訳）産業図書, 1985.

[6] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, R. Harshman: Indexing by Latent Semantic Analysis. Journal of the Society for Information Science, 41(6), 391-407, 1990.

[7] T. Hofmann: Probabilistic Latent Semantic Analysis. Proc. Uncertainty in Artificial Intelligence, 1999.

[8] 渡辺慧：知るということ, 認知科学選書, 東京大学出版会, 1986.

[9] 森本浩一：ディビッドソン, NHK 出版（シリーズ・哲学のエッセンス）, 2004.

[10]ディビッドソン：行為と出来事,（服部裕幸・柴田正良訳）勁草書房, 1990.

[11] Miller, G. A.: The science of word, Scientific American Library (1991). ; ことばの科学,（無藤隆ほか訳）東京化学同人, 1997.

[12] 柏野牧夫：音声知覚の運動理論をめぐって, 日本音響学会誌, Vol. 62, No. 5, pp. 391 – 396, (2006).

[13] Jacobson, R., Fant, G., and Halle, H.: Preliminaries to speech analysis – The distinctive features and their correlates, MIT Acoust. Lab. Tech. Rep., No. 3, 1952.

[14] 比企静雄編：音声情報処理, 東京大学出版会, 1973.

[15] Huda. M. N., Ghulam. M., Fuhuda, T., Katsurada, K., and Nitta, T.: Canonicalization of feature parameters for robust speech recognition based on distinctive phonetic feature (DPF) vectors, IEICE Trans. on Info. and Systems, E91-D(3), pp.488-498, (2008).

[16] 栗田, 小林, 三島,“PARCOR 画像の高次局所自己相関特徴を用いた背景変化および平行移動に強いジェスチャー認識,”信学技報, PRMU96-213 pp.159-164, 1997.

[17] 石井, 佐藤,“オンライン EM アルゴリズムによる動的な関数近似,”信学技報, NLP97-142, pp43-50, 1998.

[18] 田口 亮, 木村 優志, 小玉 智志, 篠原 修二, 入部 百合絵, 桂田 浩一, 新田 恒雄："幼児の学習バイアスを利用したエージェントによる語意学習の効率化", 人工知能学会論文誌 Vol.22, No.4, pp.444-453 (2007-7).

[19] 篠原 修二, 田口 亮, 橋本 敬, 桂田 浩一, 新田 恒雄："幼児エージェントにおけるバイアスの形成と言語の構造化", 情報処理学会：数理モデル化と応用研究会論文誌 Vol.48, No.SIG2(TOM16), pp.125-146 (2007-2).

[20] 篠原 修二, 田口 亮, 桂田 浩一, 新田 恒雄："語彙学習エージェントにおけるバイアスの自律調整について", 人工知能学会誌 Vol.22, No.2, pp.103-114 (2007-3).

[21] 篠原 修二, 田口 亮, 桂田 浩一, 新田 恒雄："因果性に基づく信念形成モデルと N 本腕バンディット問題への適用", 人工知能学会誌 Vol.22, No.1, pp.58-68 (2007-1).

[22] 田口 亮, 篠原 修二, 桂田 浩一, 入部 百合絵, 新田 恒雄,“エージェントによる語意学習効率化のための対話戦略獲得”, ヒューマンインタフェース学会論文誌, Vol.8, No.3, pp.71-82 , (2006-8).

[23] 岩橋直人：ロボットによる言語獲得, 人工知能学会論文誌 Vol.18, No.1a, pp.1-10 (2003).

[24] Tomasello, M.: The Cultural Origins of Human Cognition, Harvard University Press, 1999. (大堀壽夫他訳：心とことばの起源を探る, 勁草書房, 2006)

[25] 柴田義松：ヴィゴツキー入門, 子供の未来社, 2006

[26] ヴィゴツキー: 新訳版 思考と言語,（柴田義松訳）新読書社, 2001