

# Sémiologie documentaire et coordination dans l'analyse de corpus

*Transactions coopératives documédiatisées  
et recherche ouverte d'information avec des  
thésaurus, ontologies, floksonomies*

**Manuel Zacklad**

**Université de Technologie de Troyes**

Equipe **Tech-CICO** (Technologie de la Coopération pour l'Innovation et  
le Changement Organisationnel)

UTT ICD/Tech-CICO - FRE CNRS 2848

# Recherche d'information

- En matière de recherche d'information la généralisation du web évoque surtout les moteurs de recherche et les systèmes d'indexation automatiques sur lesquels ils sont basés.
- Mais elle a aussi vu récemment apparaître de nouveaux systèmes d'indexation « manuels », index interprétatifs, constitués par des groupes de professionnels ou d'utilisateurs
- Il constituent des sémiologies externes ou sémiotiques connotatives plus ou moins structurées (pour utiliser l'expression de Hjelmslev, Badir 2000), dans notre contexte des **sémiologies documentaires**

# Nouveaux développements

- Aux thésaurus utilisés par les bibliothécaires sont venus s'ajouter les ontologies dans le champ du Web Sémantique et plus récemment les Folksonomies dans les portails documentaires relevant du Web 2.0.
  - Thésaurus
  - Ontologies
  - Folksonomie (peuplonomie)

# Web Socio Sémantique

- Au sein du laboratoire Tech-CICO de l'Université de Technologie de Troyes, nous défendons depuis plusieurs années un courant original à l'intérieur du Web Sémantique que nous avons nommé le *Web Socio-Sémantique*.
- Nous considérons en particulier que la constitution des ressources terminologiques appelées à constituer des sémiologies documentaires (sémiotiques connotatives) pour les documents du Web dans un contexte intranet ou internet dépend toujours des transactions coopératives entre les communautés d'utilisateurs impliquées dans l'exploitation, le classement et la recherche de ces documents.

# Ontologies sémiotiques

- Dans l'approche du Web Socio Sémantique, la sémantique est donc pour nous intrinsèquement dépendante des interactions entre des communautés d'acteurs, des finalités de ces interactions et des problèmes de coordination qu'ils sont susceptibles de rencontrer.
- Dans ce contexte, nous définissons des « ontologies sémiotiques » (Zacklad 2005) en les différenciant des « ontologies computationnelles » du Web Sémantique, mais pas encore des thésaurus (Caussanel 2003)
- Ontologie sémiotique = un type particulier de sémiologie documentaire = un type particulier d'index interprétatif

# Recherche Ouverte d'Information

- Le Web Socio Sémantique est lui-même associé à une vision originale de la recherche d'information dite *Recherche Ouverte d'Information* (ROI, Zacklad 2005, Zaher et al. 2006),
- La finalité principale de la recherche n'est pas la récupération ou l'accès à un document ou à un ensemble de documents pertinents par rapport à une interrogation déjà formulée, mais l'assistance à un utilisateur engagé dans une démarche d'enquête à travers laquelle il sera amené à ***préciser simultanément les termes de sa demande et le corpus documentaire susceptible de contribuer à y répondre.***
- A opposer à la Recherche d'Information (RI) « fermée » qui adresse essentiellement des questions liées à d'accès.

# ROI

- Cette vision de la ROI est basée sur le développement de plateformes multifonctionnelles qui associent consultation d'annuaire (sémilogies externe), l'utilisation de moteurs de recherche, l'accès à des ressources en ligne et des échanges communautaires[1]
  - Le résultat de la ROI est constitué à la fois par un corpus de documents ressources et par un ensemble de termes[2] inter-reliés assignant à ces documents un rôle fonctionnel par rapport à un processus d'enquête en cours.
- 
- [1] On peut penser au plateformes commerciales de type Amazon mais également au développement actuels des moteurs de recherche qui combinent différentes formes d'indexation.
  - [2] Ces termes sont eux-mêmes partiellement insérés dans un système de termes décrits dans un autre document définissant la sémiologie interne associée à un domaine générique d'enquête.

# Qualité de la ROI

- L'évaluation de la pertinence/qualité de la ROI devrait donc se faire en considérant à la fois
- (1) la cohérence interne des relations entre les documents du corpus sélectionné,
- (2) la cohérence interne des relations entre les termes constituant la sous-sémiologie élaborées dans le cadre l'enquête en cours (les termes utilisés pour formuler la demande à la fois extraits d'un système existant et générés par l'enquêteur) et enfin
- (3), la cohérence des relations fonctionnelles entre les termes de la sous-sémiologie et les documents du corpus documentaire.

# RIO et recherche en SHS

- Cohérence des données : la cohérence interne des relations entre les documents du corpus sélectionné,
- Cohérence du cadre interprétatif du chercheur : la cohérence interne des relations entre les termes constituant la sous-sémiologie élaborée dans le cadre l'enquête en cours
- Cohérence externe du cadre interprétatif : la cohérence des relations fonctionnelles entre les termes de la sous-sémiologie et les documents du corpus documentaire.

# Les sémiologies chez Hjelmslev (in Badir)

		<i>analyse du plan de l'expression</i>	<i>analyse du plan du contenu</i>
<i>sémiologies internes</i>	<i>du point de vue paradigmatic</i>	phonologie	lexicologie
	<i>point de vue syntagmatic</i>	« morphologie »	grammaire
<i>sémiologies externes</i>	<i>paradigme des connotateurs historiques et géographiques</i>	phonologie historique et dialectale	lexicologie historique et dialectologie
		grammaire comparée et historique	
	<i>paradigme des connotateurs sociaux</i>	sociolinguistique, linguistique de la langue écrite	
	<i>paradigme des connotateurs psychiques</i>	pédolinguistique, psycholinguistique, analyse des troubles du langage	
	<i>paradigme des connotateurs culturels</i>	rhétorique, stylistique, narratologie	
	<i>métasémiologies internes</i>	phonétique	sémantique
	<i>métasémiologies externes</i>	physique et physiologie du son	interprétations extrinsèques

# Comparaison des index interprétatifs

- Sur la base de cette vision de la recherche d'information nous comparerons les dispositifs d'indexation « interprétatifs » (manuels) des ressources documentaires, notamment numérisées.
- Selon nous, la comparaison des avantages respectifs et des complémentarités entre ces systèmes (vocabulaire contrôlé, thésaurus, ontologies formelle et sémiotique, folksonomie...) ne doit pas se faire uniquement en considérant les propriétés structurelles des méta-sémiotiques (Hjelmslev 1966) sur lesquelles ils s'appuient mais en tenant compte de leur rôle dans la conduite des activités de ROI et des modalités de leur constitution et de leur maintenance.

# Approche englobante

- Cela implique de tenir compte des efforts entrepris pour constituer et maintenir les sémiologies documentaire de manière manuelle du point de vue des compétences acteurs et de la dynamique sociale associée à ces processus de constitution (Caussanel et al. 2003, Lejeune 2006).
- La comparaison devra aussi tenir compte du rôle de ces systèmes dans les plateformes plurifonctionnelles de ROI qui combinent aujourd'hui index automatiques et index interprétatifs.

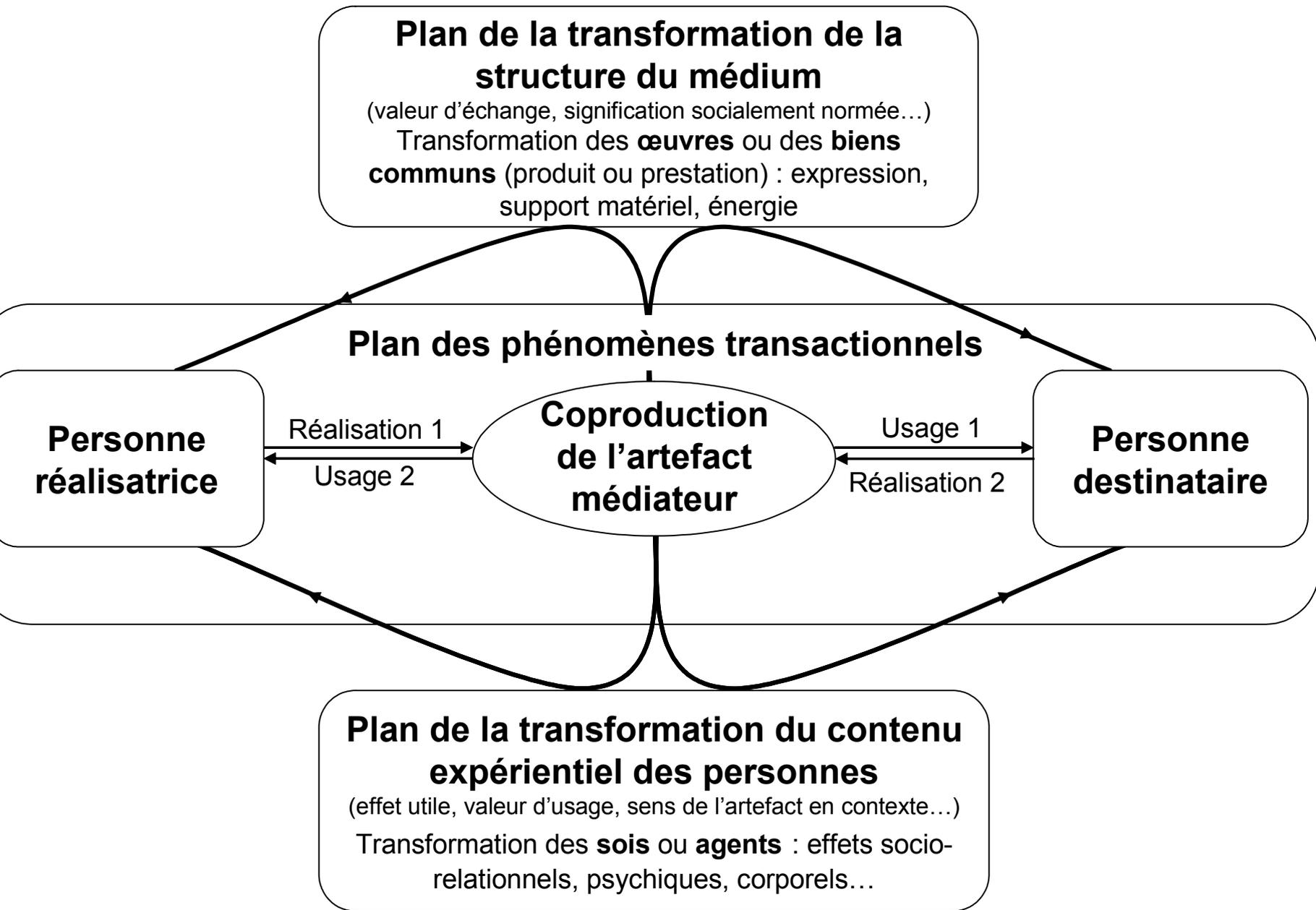
Un cadre théorique pour la  
compréhension du statut des  
sémiologies documentaires dans  
l'activité collective

# Plan de la deuxième partie

- Le cadre d'analyse des transactions coopératives : flux transactionnel et niveau de coordination
- Le statut des artefacts documentaires et la documentarisation
- Forme de coordination et structure des sémiologies interprétatives

# Transactions dans la théorie transactionnelle de l'action

- Nous empruntons à Bentley et Dewey (1949) qui le distingue de celui d'interaction
- Chaque transaction est l'occasion d'une double transformation d'un artefact (expressif ou matériel) et des personnes dont les contours sont redéfinis
- La personne physique en s'engageant dans une transaction devient un acteur **individuel** ou **collectif** (représenté par un porte parole) ou la même personne engagée dans une forme « dialogue » avec elle-même.



Schématisation des composantes d'une transaction

# Moments de la transaction d'un point de vue individuel et social

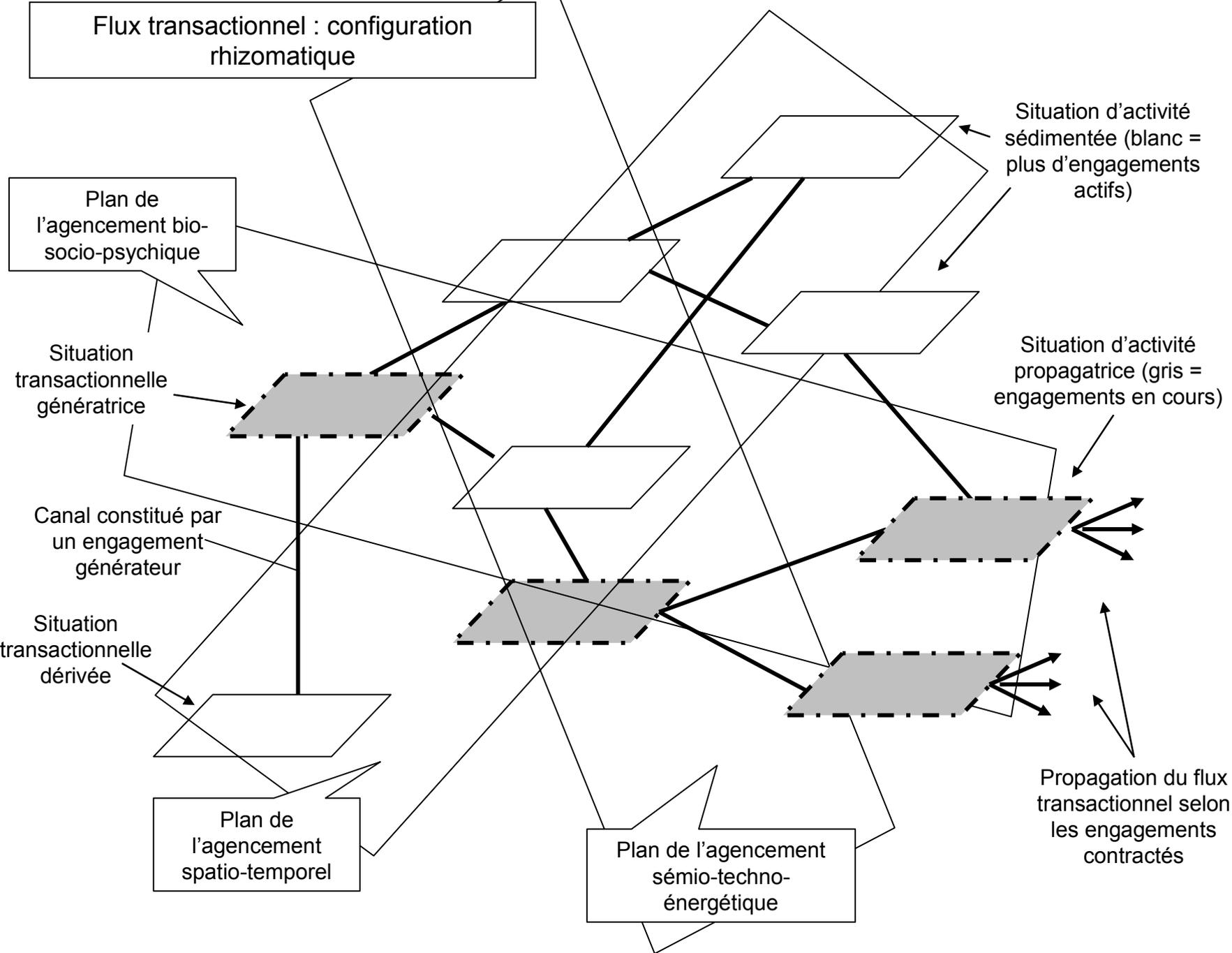
- Les quatre temps de la transaction coopérative (Greimas) :
  - **Virtualisation** (le réalisateur intègre le projet du bénéficiaire), ou « manipulation »
  - **Acquisition de compétences** (acquisition de moyens ressources)
  - **Réalisation de la performance** (transformation, distribution, réception, appropriation d'un médium)
  - **Sanction** (rétribution symbolique, matérielle, financière)

# Transaction de l'économie néoclassique vs transaction coopérative

- Dans le vocabulaire traditionnel de l'économie la transaction correspond aux relations qui se déroulent sur un marché
  - Modèle néoclassique du marché : anonymat des protagonistes, indépendance des préférences, absence de relation interpersonnelle entre les protagonistes, caractère ponctuel de la relation marchande, faible incertitude
- La transaction coopérative s'oppose à priori en tout point à ce sens restreint
- Elle inclut les temps d'élaboration de la solution, d'usage et de consommation

# Situation d'activité et flux transactionnel

- Les transactions peuvent être décomposées en de nombreuses sous-transactions
- La transaction progresse toujours dans le cadre de **situations d'activité** durant lesquelles les acteurs transforment activement les artefacts et les personnes
- La réalisation d'une transaction peut être interrompue et s'étaler sur de longues périodes de temps, à l'occasion de plusieurs situations d'activité
- Nous considérerons que les transactions correspondent à des flux qui s'écoulent à des vitesses variables en s'entrecroisant à divers endroits.
- Toute situation d'activité effective peut ainsi être analysée en fonction de sa « généalogie » au sein d'un flux transactionnel



# Configuration rhizomatique et machine transactionnelle

- L'écoulement libre des flux transactionnels génère des configurations rhizomatiques (Deleuze & Guattari 1980) dont les nœuds sont les situations d'activité effectives.
- Dans d'autres cas les flux transactionnels sont régis par des règles formelles et se répètent de manière cyclique : on parle alors de configuration machinique ou de **machine transactionnelle**
- Dans la machine transactionnelle, les situations d'activités effectives sont autant que possible agencées selon des règles qui reflètent la hiérarchie sociale et l'organisation systématique de l'activité.
- Dans les deux configurations, un enjeu majeur, pour lequel les TIC seront sollicitées, est de faciliter la circulation entre ces situations d'activité

# Coordination et régulation des flux transactionnels

# La coordination

- La coordination correspond pour nous à l'ensemble des artefacts et des méthodes permettant d'assurer une continuité entre les situations d'activités appartenant à un même flux transactionnel.
- Les problèmes de coordination se posent à l'intérieur des situations d'activité (coordination intra-situationnelle), mais également entre des situations d'activité différentes au sein d'un flux transactionnel (coordination inter-situationnelle)

# Trois niveaux de coordination

- L'accès aux situations d'activité
- La standardisation de la transaction primaire
- Le recours à des ressources régulatrices

# L'accès aux situations et à leurs composantes

- Synchronisation temporelle, spatiale, accès à l'espace, aux artefacts, aux personnes...
- Problématique de la transformation du support des artefacts pour faciliter leur « circulation » :
  - Dimension expressive dominante :
    - passage d'un support éphémère à un support pérenne  
-> recours au documents
  - Dimension physique dominante :
    - Transformation des qualités physiques -> conservation des produits
    - Aménagement d'un local

# Dans l'exemple de la fourniture « d'accès internet »

- Exemple : magasin pour la rencontre entre le client et le fournisseur, lignes téléphoniques, infrastructures du web, document numérique ou papier...

# Coordination par la standardisation préalable de la transaction primaire

- Distinction entre transaction **primaire** et transaction **secondaire** ou « régulatrice »
  - Distinction entre travail coopératif et travail d'articulation du travail coopération (Schmidt et Simone)
  - Notion de travail d'organisation (De Terssac)
- Deux principales modalités
  - Préalablement à la réalisation de la transaction primaire pour régler, à l'avance, les interdépendances
  - Pendant le déroulement de la transaction primaire

# Modalités de la standardisation

- Ce traitement anticipé des interdépendances a pour conséquence une standardisation du déroulement de la transaction primaire et de ses résultats
- Standardisation des artefacts
  - Matériel et énergétique (interopérabilité des composants et des mouvements)
  - Expressif : cohérence « logique » dans un formulaire
- Standardisation des personnes et de leurs relations
  - Corporel (synchronisation posturale et motrice)
  - Psychique (habitudes, automatismes, attitudes...)
  - Social (normativité des rôles)

# Dans l'exemple de la fourniture d'accès

- Au niveau des artefacts
  - Standardisation des protocoles techniques permettant l'accès au réseau
  - Standardisation des fonctionnalités
  - Standardisation des bases de données
- Au niveau des personnes
  - Standardisation des relations de conseil et d'assistance entre le fournisseur et le client
  - Standardisation des relation contractuelles

# Coordination par le recours à des transactions et des ressources régulatrices en situation

- Elaborer des dispositifs qui visent à faciliter le travail de régulation par les personnes engagées dans la transaction primaire
- Mobiliser des attitudes et compétences qui permettent cette adaptation
- Coopération « active » vs « passive »

# Ressources régulatrices en situation

- **Artefact**
  - Matériel : plurifonctionnalité des composants
  - Energétique : diversité des procédures associées à une prestation
  - Expressif : potentialisation sémiotique du contenu, dimension artistique...
- **Personnes**
  - Corporel : nouvelles formes de synchronisation motrice
  - Psychique : capacité d'interprétation et d'apprentissage
  - Social : ouverture vers l'intersubjectivité des tiers

# Dans le cas de l'exemple

- Artefact

- Matériel : diversité de fonctionnalités en mesure de s'adapter au besoin du client
- Expressif :
  - Richesse de la documentation technique
  - Qualité des dialogues

- Personnes

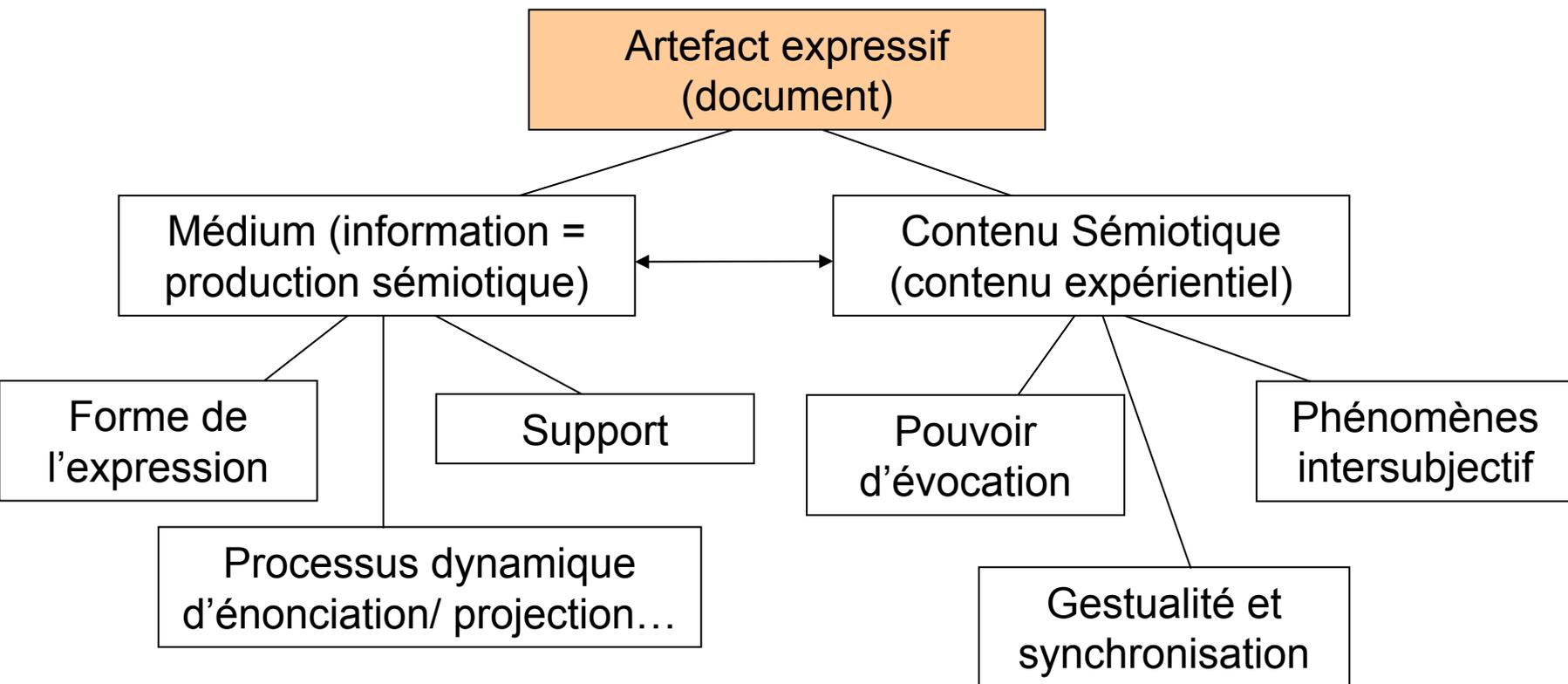
- Psychique : capacité d'apprentissage du client et du prestataire
- Social : capacité de se « mettre à la place » de l'interlocuteur

# Théorie du document

# Une théorisation documentaire des TIC

- Avantage de l'approche en termes de « document »
  - Décentration vis à vis de la vision technique pour se focaliser sur les usages et les pratiques qui sont en partie en continuité avec ceux qui préexistaient à l'informatisation
  - Permet de rendre compte des derniers développements du Web marqués par la généralisation à grande échelle de transactions médiatisées par des documents textuels et multimédia (usage des blogs et des wiki, Web 2.0),

# Dimensions de la structure du médium et du contenu expérientiel dans un artefact expressif



# Documentarisation

- Deux types de supports matériels :
  - *Les supports éphémères : voix, geste, sable sur la plage...*
  - *Supports pérennes : papier, pierre, numérique...*
- Il existe deux méthodes permettant de véhiculer une production sémiotique sur un support pérenne :
  - Transcription : le support de signes graphiques dans des logiques de figuration ou d'équivalence phonétique, par exemple
  - L'enregistrement : via des dispositifs le plus souvent aujourd'hui électroniques

# Investissement documentaire de la transcription ou de l'enregistrement

- Transcription et enregistrement sont des conditions nécessaires mais non suffisantes de la documentarisation
- Une prise de notes ou un enregistrement ponctuel visant à faciliter une activité sémiotique dans une situation transactionnelle donnée pourront avoir une utilité locale certaine.
- S'ils ne font pas l'objet d'un *investissement documentaire* permettant de soutenir les processus de documentarisation, ces supports ne pourront pas être *ré exploités* dans des situations transactionnelles différées et distribuées

# Définition de la documentarisation

- Consiste à transcrire ou à enregistrer les contenus sémiotiques sur des supports pérennes puis à doter ces supports d'attributs spécifiques constituant une **sémiotique connotative** (sémiologie externe) permettant :
  - de faciliter leur articulation « interne » : transformation du support pour permettre l'accès aux différentes parties du contenu sémiotique selon un découpage éditorial donné (titre de chapitre, numérotation, table des matières, métadonnées...).
  - de faciliter leur articulation « externe » : transformation du support pour permettre son rangement cohérent dans des collections ou des archives (titre, cotes du document, mots-clefs...).

# Définition du document

- Document : *production sémiotique transcrite ou enregistrée sur un support pérenne qui est équipée d'attributs spécifiques visant à faciliter les pratiques liées à son exploitation ultérieure (ré-exploitation) dans le cadre de la préservation de transactions communicationnelles réparties*

# Une définition en ligne avec les fondamentaux de la documentologie

- La catégorie *document* inclut aussi des documents multimédia
  - œuvre graphique, photographie, cassette audio et vidéo, CD, DVD, jeux vidéo, etc., voire des objets manufacturés à vocation artistique, des objets liturgiques, (cf. Briet 1951, Buckland 1997).
  - Rôle essentiel de « preuve »
  - Exemple de l'antilope de Briet

**Conclusion : retour sur les  
modalités d'indexation et de  
recherche ouverte d'information**

# Sémiologie documentaire et coordination

- Les sémiologies documentaires correspondent à des méthodes de documentarisation dans un contexte de transactions coopératives hautement distribuées
- Chaque type de sémiologie ou d'index interprétatif renvoie à des mécanismes de coordination qui sont eux-mêmes liés à des configurations organisationnelles (documents privés) ou à des principes de gouvernance (documents publics)

# Exemples et hypothèses

- Le mode thésaurus : est liée à l'existence d'une expertise liée au corpus, un corps de représentants experts assurant la médiation avec le « public »
- Le mode ontologie computationnelle (web sémantique) : même phénomène sur la base d'une expertise philosophico/scientifique basés sur une sémiotique transcendantale
- Le mode floksonomique (peuplonomique) « pur » : participation populaire sans représentation ni structuration ni des collectifs ni des connaissances
- Le mode ontologie sémiotique : effort pour articuler structuration du collectif et du domaine (points de vue multiples et complémentaires)

# Transposition dans le contexte de la recherche scientifique SHS

- Approche descendante : guidée par la théorie (ontologies a priori)
- Approche ascendante pure : guidée par le corpus « grounded theory », approche peuplonomique
- Principe d'analyse alternant les deux démarches: ontologies sémiotiques multipoints de vue
- Adaptation particulière aux études pluridisciplinaires de corpus

**Merci**