

Normes et pratiques exemplaires en technologies langagières

Language Technology Standards and Best Practices

Auteur : Marie-Hélène Lebeault, consultante en communication

Introduction

Le présent document vise à fournir une information essentielle sur les méthodes de travail et l'état de l'art en matière de technologies langagières. Il ne fait pas office de prescription ou de réglementation en la matière. Les normes recensées émanent autant de l'industrie privée (Microsoft Corp.) que d'organismes publics (W3 Consortium). Le rôle de ces normes est absolument fondamental dans l'industrie : permettre l'échange des connaissances et l'émulation (contexte de recherche) et la compatibilité des solutions (pour le développement). Nous avons voulu fournir ici une « référence rapide » qui permettra aux artisans des technologies de la langue de prendre connaissance des jalons posés dans leur spécialité respective. Par souci de commodité, le document est présenté sous forme de tableaux bilingues.

1 – Multimédia

Normes et standards du multimédia

		HTML	SGML	Hytime	MHEG-5	ODA/Hyper ODA	MPEG-4 et 7
Élément Multimédia/Multimedia element	Texte Hypertexte	Oui (10/10)	Oui (10/10)	Oui (10/10)	Oui (8/10)	Oui (10/10)	-
	Image fixe	Oui (10/10)	Oui (10/10)	Oui (10/10)	Oui (10/10)	Oui (10/10)	-
	Image animée	Oui (3/10)	Non	Oui (10/10)	Oui (10/10)	Oui (5/10)	Oui (10/10)
	Son	Oui (5/10)	Non	Oui (10/10)	Oui (10/10)	Oui (5/10)	Oui (10/10)
	3D	Oui (5/10)	Non, Pas de support audio,	Oui (10/10)	Oui (10/10)	Non (texte et image pour ODA, tout pour HyperOda)	Oui (?/10)

			vidéo ou graphique				
Communication Réseau/Network communication	Oui (10/10)	Oui (5/10) Références croisées seulement à l'intérieur d'un même document	Oui (10/10)	Oui (10/10)	Oui (2/10) ODA est incapable de présenter plusieurs exemplaires d'un même document Hyper ODA prend en compte des hyperliens	Oui	
Structuration du document/Document structure	Oui (8/10)	Oui (10/10)	Oui (10/10)	Oui (10/10)	Oui (2/10) Faible, basée sur des rectangles	?	
Relation Temporelle/Time relationship	Oui (5/10) permet d'incorporer l'information multimédia dans les documents	Non, Ne supporte pas les données dépendantes du temps	Oui (10/10) Prise en compte d'éléments multimédias synchrones et synchrones	Oui (10/10) Permet le temps réel sur WAN (Interactivité), synchronisation multimédia	Non (2/10). Permet d'incorporer l'information multimédia dans les documents ; supporte mal l'interactivité	Oui	
Portabilité/Portability	Oui (10/10)	Oui (10/10)	Oui (10/10) Identification, représentation et structuration uniforme des données multimédias	Oui (8/10) Encodage indépendant du système	Oui (5/10) Structures génériques	Oui	
Évolutivité/Scalability	Oui (7/10)	Oui (10/10)	Oui (10/10)	Oui (10/10)	Oui (10/10) Différencie la structure du contenu	Oui	
Présence d'outils de création, contrôle/Integration of creative/control tools	Oui (10/10)	Oui (10/10)	Oui (4/10) en cours	Oui (6/10) en cours	Oui (4/10)	-	
Marché/Market	Oui (10/10) World Wide Web	Oui (7/10) Surtout celui de l'information textuelle	Oui (4/10) Segments spécifiques	Oui (5/10)	Oui (2/10) Journaux, livres...	-	
Norme et standard associé/Related standard and norm	Java, Javascript Et VRML			DSM-CC DAVIC			

2 – Interfaces parole/Speech interfaces

Français	Norme/Standard	English
La norme SALT est conçue pour faciliter l'accès multimodal et téléphonique aux informations, applications et services Web, à partir de téléphones et de systèmes mobiles sans fil.	SALT Speech Application Language Tags	This specification enables multimodal and telephony-enabled access to information, applications, and Web services from PCs, telephones, tablet PCs, and wireless personal digital assistants (PDAs).
Sous-titrage	SAMI Synchronized Accessible Media Interchange (Microsoft)	Captioning
Création de présentations audiovisuelles interactives, généralement utilisée pour les « médias enrichis ».	SMIL Synchronized multimedia integration language (W3C)	Authoring of interactive audiovisual presentations, typically used for "rich media".

3 – Traitement du langage naturel /Natural Language Processing

Français	Norme/Standard	English
La Text Encoding Initiative (TEI) est un projet international qui vise la mise au point d'un ensemble de normes pour la préparation et l'échange de textes électroniques.	TEI Text Encoding Initiative	"TEI" is short for "Text Encoding Initiative." The TEI was founded in 1987 to develop guidelines for encoding machine-readable texts of interest in the humanities and social sciences.
La norme CES, fournit un ensemble de balises spécifiques au codage des corpus de textes pour les besoins de l'ingénierie	CES ##Corpus Encoding Standard	The CES is designed to be optimally suited for use in language engineering research and applications, to serve as a set of encoding

linguistique, et leur sémantique précise dans le contexte des corpus.		standards for corpus-based work in natural language processing applications.
	Organismes/Organisations	
Organisme qui promeut l'élaboration de ressources pour le traitement de la parole et l'évaluation des technologies de la parole	COCOSDA Comittee for Co-ordination and Standardisation of Speech Databases	Organisation supporting the development of speech processing resources and assessment of speech technologies.
Groupe d'agences qui encouragent la recherche et le développement technologique par la création et le partage de ressources linguistiques (données, outils et normes).	LDC Linguistic Data Consortium	Consortium that supports language-related education, research and technology development by creating and sharing linguistic resources: data, tools and standards.
Repérage, diffusion, collecte, validation, standardisation, améliorations des ressources linguistiques. Promeut aussi la recherche scientifique relative aux ressources linguistiques et à l'évaluation des systèmes.	ELRA European Language Resources Association	Identification, distribution, collection, validation, standardisation, improvement, promoting the production of language resources. Also supports scientific research related to language resources and system evaluation.

4 – Reconnaissance et synthèse de la parole/Speech Recognition and synthesis

Français	Norme/Standard	English
La norme SAPI est un API conçu par Microsoft pour utiliser la reconnaissance vocale et la synthèse vocale à partir des applications Windows.	SAPI Speech Application Programming Interface	SAPI is an API developed by Microsoft to allow the use of Speech Recognition and Speech Synthesis within Windows applications.

<p>La norme JSAPI fournit une couche d'interfaçage entre une application JAVA et un synthétiseur ou un moteur de reconnaissance de voix.</p>	<p style="text-align: center;">JSAPI Java Speech Application Programming Interface</p>	<p>JSAPI allows Java applications to incorporate speech technology into their user interfaces. It defines a cross-platform API to support command and control recognizers, dictation systems and speech synthesizers.</p>
<p>Il s'agit de normes secondaires. JSML (beta) définit les balises pour le codage du texte entré dans un synthétiseur. JSGF fournit les balises grammaticales pour la reconnaissance vocale.</p>	<p style="text-align: center;">JSML and JSGF Java Speech API Markup Language and Java Speech API Grammar Format</p>	<p>They are companion specifications to JSAPI. JSML (currently in beta) defines a standard text format for marking up text for input to a speech synthesizer. JSGF version 1.0 defines a standard text format for providing a grammar to a speech recognizer.</p>
<p>Les spécifications de la plate-forme interface vocale VoiceXML fournissent les directives pour créer des dialogues vocaux entre un téléphone et un ordinateur.</p>	<p style="text-align: center;">VoiceXML</p>	<p>VoiceXML is a markup language that defines a common format for dialog between a caller and a computer system with speech content.</p>
<p>SSML est une nouvelle norme qui permet de contrôler un moteur de synthèse de la parole.</p>	<p style="text-align: center;">SSML Speech Synthesis Markup Language</p>	<p>SSML is an emerging standard that defines how portions of a text should be spoken by any TTS engine.</p>
<p>Le protocole de communication MRCP permet aux serveurs vocaux d'offrir des services vocaux, tels que la synthèse et la reconnaissance vocale, à leurs clients. Il agit comme un « contenant » mettant en œuvre les diverses technologies. On trouve MRCP v1 (RFC 4463), très répandu dans l'industrie, et MRCP v2, nouveau protocole de l'IETF, basé sur SIP/RTP.</p>	<p style="text-align: center;">MRCP Media Resource Control Protocol</p>	<p>MRCP is a communication protocol, which allows speech servers to provide various speech services (such as speech recognition and speech synthesis) to its clients. This is a "wrapper" which implements the various required technologies. There are 2 versions: MRCP v1, which is widely used in the industry, and MRCP v2 new IETF protocol, based on SIP/RTP.</p>
<p>SOAP est un protocole fondé sur XML pour l'échange d'informations en environnement</p>	<p style="text-align: center;">SOAP</p>	<p>SOAP is a lightweight protocol for exchange of information in a decentralized, distributed</p>

décentralisé. Il fait l'objet d'une recommandation de la part du W3C, et est couramment utilisé pour établir un canal de communication entre services web.	Simple Object Access Protocol	environment. It is an XML-based protocol that is not specific to speech, but is frequently used to retrieve data across a network.
La norme SRGS, une grammaire de reconnaissance vocale, permet aux auteurs de spécifier des règles qui prennent en compte des séquences de mots que les utilisateurs pourraient dire dans un contexte donné. Elle décrit les choix possibles des utilisateurs en réponse à des annonces vocales.	SRGS Speech Recognition Grammar Specification	SRGS defines syntax for representing grammars for use in speech recognition so that developers can specify the words and patterns of words to be listened for by a speech recognizer.
La norme SISR est un langage utilisé par le moteur de reconnaissance vocale pour décrire les mots et expressions reconnues.	SISR Semantic Interpretation for Speech Recognition	SISR defines the syntax and semantics of annotations to grammar rules in the Speech Recognition Grammar Specification (SRGS).
La norme PLS est un langage de spécification de la façon dont les mots sont prononcés pour les moteurs ASR et TTS.	PLS Pronunciation Lexicon Specification	PLS is designed to enable interoperable specification of pronunciation information for both ASR and TTS engines within voice browsing applications.
la spécification GRDDL définit un mécanisme pour extraire et transformer des microformats XML, qui permet de décrire les métadonnées), rouage essentiel du Web sémantique.	GRDDL Gleaning Resource Descriptions from Dialects of Languages	The GRDDL specification introduces markup for declaring that an XML document includes gleanable data and for linking to an algorithm, typically represented in XSLT, for gleaning the resource descriptions from the document.

5 – Traduction automatique ou assistée/Machine-based /Automated /Computer-assisted translation

Français	Norme/Standard	English
Le langage SGML, un langage normalisé permettant de décrire les relations entre le contenu d'un document informatique et sa structure, spécifie un système de marquage du contenu d'un document qui assure en particulier que tout document pourra être relu, quelle que soit l'évolution des systèmes et des traitements de texte.	SGML Standard Generalized Markup Language	SGML enables documents to be broken down and stored as information units. Storage of these units within a suitable repository enables collaborative authoring, life-cycle tracking, and automatic configuration of the units into document deliverables.
La norme TDT définit les protocoles de lecteur nécessaires pour transférer les données et le service de base pour la traduction des EPC dans le lecteur vers une méthode de représentation appropriée.	TDT EPC Tag Data Translation	TDT specification describes the ways to interpret machine-readable tag data. This standards contains details of the structure and elements of the machine readable markup files and provides guidance on how it validate it using a software system tag data translation.
TMX est un format de base XML permettant un échange normalisé des données contenues dans les mémoires de traduction.	TMX Translation Memory Exchange	This XML standard enables the interchange of translation memories between translation suppliers.
TBX et un format de base XML permettant un échange normalisé de données entre différentes bases de données terminologiques.	TBX Termbase Exchange	TBX is an open XML-based standard format for terminological data.
La norme SRX a pour objectif de spécifier davantage les règles de segmentation qui ont servi à créer une mémoire de traduction balisée selon TMX.	SRX Segmentation Rules Exchange	SRX is intended to enhance the TMX standard so that translation memory data that is exchanged between applications can be used more effectively.

<p>La norme GMX a pour ambition de permettre de quantifier un projet de traduction de manière standardisée.</p>	<p align="center">GMX GILT Metrics standard</p>	<p>The proposed GILT Metrics standard is tasked with quantifying the workload and quality requirements for any given GILT task. GILT stands for (Globalization, Internationalization, Localization, and Translation).</p>
<p>La format OLIF permet l'échange de ressources lexicales et terminologiques selon une approche ouverte à la polysémie</p>	<p align="center">OLIF Open Lexicon Interchange</p>	<p>OLIF is an open, XML-compliant standard for the exchange of terminological and lexical data. Although originally intended as a means for the exchange of lexical data between proprietary machine translation lexicons, it has evolved into a more general standard for terminology exchange.</p>
<p>Le format XLIFF permet de localiser directement un produit logiciel sans avoir à se soucier du format natif.</p>	<p align="center">XLIFF XML Localisation Interchange File Format</p>	<p>XLIFF is intended to provide a single interchange file format that can be understood by any localization provider.</p>
<p>Transweb permet d'identifier les appels nécessitant les services internet pour soumettre et récupérer les fichiers et messages d'un projet de localisation.</p>	<p align="center">TransWS Translation Web Services</p>	<p>TransWS specifies the calls needed to use Web services for the submission and retrieval of files and messages relating to localization projects.</p>
<p>Xml: tm repose sur le partage de mémoires de traduction à travers différentes plateformes ou logiciels et qui peuvent comporter plus d'une langues.</p>	<p align="center">xml:tm</p>	<p>xml:tm is .based on the concept of text memory which comprises author and translation memory. It is a syntax that uses the power of XML to embed additional information within the XML document itself.</p>

<p>XPDL est un standard de la Workflow Management Coalition qui permet de définir un processus métier à l'aide du langage XML. Processus métier qui sera ensuite utilisé par un moteur de workflow.</p>	<p style="text-align: center;">XPDL XML Process Definition Language</p>	<p>XPDL is a format standardized by the Workflow Management Coalition to interchange Business Process definitions between different workflow products like modeling tools and workflow engines.</p>
<p>BPEL est un langage de description des procédures d'entreprise qui est issu des langages WSLF (Web Services Flow Language) et XLANG. Il est sérialisé en XML et vise à rendre possible le <i>programming in the large</i>.</p>	<p style="text-align: center;">BPEL Business Process Execution Language</p>	<p>BPEL is a business process modeling language that is executable. The origins of BPEL can be traced to WSFL and XLANG. It is serialized in XML and aims to enable programming in the large.</p>

6 – Formation virtuelle/e-Learning

Français	Norme/Standard	English
<p>SCORM est un cadre de caractéristiques adapté à partir de sources multiples qui fournit une suite complète de capacités d'apprentissage électronique permettant l'interopérabilité, l'accessibilité et la réutilisabilité du contenu d'apprentissage sur le Web.</p>	<p>SCORM Shareable Content Object Reference Model</p>	<p>Shareable Content Object Reference Model (SCORM) is a collection of standards and specifications for web-based e-learning. It defines communications between client side content and a host system called the run-time environment (commonly a function of a learning management system).</p>
<p>Les AGR sont un ensemble de 9 normes conçues par la AICC en matière d'enseignement assisté par ordinateur.</p>	<p>AGRs AICC Guidelines and Recommendations</p>	<p>AICC Guidelines and Recommendations (AGRs) are the most widely accepted interoperability standards for computer-based and web-based training pioneered by The Aviation Industry Computer-Based Training Committee (AICC).</p>
<p>LOM est un modèle standardisé destiné à la description d'objets pédagogiques, quelle que soit la forme de ces objets. LOM est structuré de manière hiérarchique à l'aide de 78 éléments sur 3 niveaux, et peut se décliner suivant des modèles d'application correspondant à des besoins particuliers. Il existe ainsi une version française de LOM, LOMFR et une Canadienne CanCore.</p>	<p>LOM Learning Object Metadata</p>	<p>Learning Object Metadata is a data model, usually encoded in XML, used to describe a learning object and similar digital resources used to support learning. The purpose of learning object metadata is to support the reusability of learning objects, to aid discoverability, and to facilitate their interoperability, usually in the context of online learning management systems (LMS).</p>

<p>Le mandat de ce groupe de travail consiste principalement à « adapter, corriger, amender ou améliorer les aspects techniques du standard IEEE nommément appelé LOM (Learning Object Metadata) » Ces efforts ont comme principal objectif d'établir une compatibilité avec le standard IEEE LOM, mais ce ne sera pas nécessairement le seul standard ou cadre de référence abordé pour le développement du LRM.</p>	<p style="text-align: center;">LRM Learning Resource Meta-data</p>	<p>The IMS LRM specification provides an extensive Best Practice and Implementation Guide, and an XSL transform that can be used to migrate metadata instances from the older versions of the IMS LRM XML binding to the IEEE LOM XML binding.</p>
<p>Le mandat de ce groupe de travail est non pas d'étudier la création d'une nouvelle norme, mais plutôt d'enquêter sur les efforts de normalisation existants et de déterminer la désirabilité et la faisabilité de les influencer afin qu'ils soient utilisables dans un contexte d'apprentissage</p>	<p style="text-align: center;">DREL Digital Rights Expression Languages</p>	<p>The Digital Rights Expression Languages (DREL) workgroup within the IEEE Learning Technology Standards Committee is gathering requirements that a standardized DREL must meet to support learning, education, and training.</p>
<p>CMI, ou Enseignement géré par ordinateur (EGO), désigne principalement la gestion administrative d'un enseignement (inscription, suivi des apprenants, stockage des scores...). Ce terme est aujourd'hui moins utilisé au profit du terme LMS (Learning Management System).</p>	<p style="text-align: center;">CMI Computer Managed Instruction</p>	<p>Computer-managed instruction is an instructional strategy whereby the computer is used to provide learning objectives, learning resources, and assessment of learner performance. Computer-managed instruction (CMI) aids the instructor in instructional management without actually doing the teaching.</p>

<p>Le LIP permet d'établir un profil personnel de préférences pour une présentation sur mesure du contenu d'apprentissage. Soit comment afficher le contenu, les appareils d'entrée préférés ou requis et les contenus alternatifs préférés et outils de soutien.</p>	<p>IMS LIP Learner Information Package</p>	<p>The IMS Learner Information Package (IMS LIP) specification addresses the interoperability of internet-based Learner Information systems with other systems that support the Internet learning environment. The intent of the specification is to define a set of packages that can be used to import data into and extract data from an IMS compliant Learner Information server.</p>
<p>La spécification AccessForAll porte sur la description des contenus d'apprentissage.</p>	<p>IMS AccessForAll Meta-data</p>	<p>The AccessForAll Meta-data specification is intended to make it possible to identify resources that match a user's stated preferences or needs.</p>
<p>RDCEO est une spécification qui permet de créer une compréhension partagée des compétences telles qu'elles apparaissent dans un plan d'apprentissage ou dans un plan de carrière, dans des préalables ou dans des résultats d'apprentissage. Cette spécification propose un modèle d'information pour décrire, référencer et échanger des définitions de compétences ou d'objectifs pédagogiques.</p>	<p>IMS RDCEO Reusable Definition of Competency or Educational Objective</p>	<p>This specification defines an information model for describing, referencing, and exchanging definitions of competencies, primarily in the context of online and distributed learning. It enables interoperability among learning systems that deal with competency information by providing a means for them to refer to common definitions with common meanings</p>
<p>Le portfolio est un dossier personnel qui permet de présenter les réalisations d'un apprenant, de garder les traces de son apprentissage et de présenter ses résultats.</p>	<p>ePortfolio</p>	<p>The IMS ePortfolio specification was created to make ePortfolios interoperable across different systems and institutions.</p>

<p>Le IMS Learning Design est une spécification qui s'inscrit dans un vaste mouvement international de normalisation pour favoriser la réutilisation des ressources d'apprentissage et l'interopérabilité des systèmes d'apprentissage.</p>	<p style="text-align: center;">LD Learning Design</p>	<p>The IMS Learning Design specification supports the use of a wide range of pedagogies in online learning. Rather than attempting to capture the specifics of many pedagogies, it does this by providing a generic and flexible language. This language is designed to enable many different pedagogies to be expressed..</p>
<p>IMS/QTI (IMS Question & Test Interoperability Specification) est une spécification de description de contenu concernant les exercices interactifs. Elle est proposée par IMS Global learning consortium. et permet la séparation des données (les questions, les bonnes réponses etc.) de la forme (affichage) et du traitement (calcul des scores etc.). Ces spécifications se basent sur un schéma XML public</p>	<p style="text-align: center;">QTI Question & Test Interoperability</p>	<p>The QTI specification describes a data model for the representation of question and test data and their corresponding results reports. It enables the exchange of this item, test and results data between authoring tools, item banks, test constructional tools, learning systems and assessment delivery systems.</p>
<p>Cette norme concerne l'interopérabilité des listes de ressources. Elle détaille le moyen d'échanger des métadonnées structurées entre des systèmes qui stockent et exposent des ressources pour créer des listes de ressources, et ceux qui recueillent et organisent ces mêmes listes à des fins d'enseignement et de formation. La liste des livres recommandés en est un exemple type.</p>	<p style="text-align: center;">RLI Resource List Interoperability</p>	<p>The Resource List Interoperability (RLI) specification details how structured meta-data can be exchanged between systems that store and expose resources for the purpose of creating resource lists and those that gather and organize those Resource Lists for educational or training purposes. A typical example of such a resource list is a reading list.</p>

<p>IMS Simple Sequencing est une spécification utilisée pour décrire un cheminement à travers un ensemble d'activités pédagogiques. IMS SS décrit un ordre dans lequel les activités pédagogiques seront présentées à l'apprenant et les conditions pour lesquelles les ressources sont sélectionnées, délivrées ou outre passées durant l'apprentissage.</p>	<p>Simple Sequencing</p>	<p>The IMS Simple Sequencing Specification defines a method for representing the intended behavior of an authored learning experience such that any learning technology system (LTS) can sequence discrete learning activities in a consistent way.</p>
<p>Le IMS Vocabulary Definition Exchange est une spécification qui définit une grammaire pour l'échange de termes. Plus spécifiquement, VDEX définit une grammaire pour l'échange de listes informatisées de termes de même que des compléments d'information qui permettront à des êtres humains de comprendre leur sens et leur portée.</p>	<p>VDEX Vocabulary Definition Exchange</p>	<p>VDEX defines a grammar for the exchange of simple machine-readable lists of values, or terms, together with information that may aid a human being in understanding the meaning or applicability of the various terms.</p>
<p>IMS Content Packaging propose un modèle d'information pour décrire l'échange de ressources pédagogiques sur Internet. Il propose un ensemble de structures pour la description des différentes composantes d'une ressource et l'interopérabilité entre les outils de création de ressources pédagogiques, les système de gestion de ressources pédagogiques et les systèmes d'exécution de ressources pédagogiques.</p>	<p>CP IMS Content Packaging</p>	<p>The IMS Content Packaging (CP) Information Model describes data structures that are used to provide interoperability of Internet based content with content creation tools, learning management systems (LMS), and run time environments. The objective of the IMS CP Information Model is to define a standardized set of structures that can be used to exchange content.</p>

7 – Métadonnées audio/Audio metadata

Français	Norme/Standard	English
Format de données « cart chunk » - automatisation radio et catalogage	AES46	"Cart Chunk data" format – Raudio automation and cataloging

Conclusion

L'importance des normes comme moyen de communication et d'échange est inestimable. À telle enseigne que des entreprises ont tenté de déposer un droit d'auteur sur certaines normes « de fait ». Le débat est en cours aux É.-U. Vous pouvez en apprendre davantage sur ce sujet en consultant le magazine *Communication* de l'Association for Computer Machinery (ACM).

Standards are absolutely essential as a mean of communication and information exchange, to the point where some corporations have tried to file a copyright on certain *de facto* standards. More on the subject in *Communication* magazine of the ACM (Association for Computer Machinery).