

Annexe H

Histoire de l'unification han

Dès les premières normes nationales comprenant les caractères han, on assiste à des efforts pour créer un code unifié reprenant ces caractères. Le Code de caractères chinois pour l'échange d'information (CCCI) a été conçu en 1980 à Taïwan, il reprend les caractères utilisés en Chine, à Formose et au Japon. Les États-Unis l'ont adopté, sous une forme quelque peu modifiée, sous le nom d'ANSI Z39.64-1989 ou encore, dans le domaine bibliographique, du Code de caractères d'Asie orientale (EACC).

Le jeu de caractère Han de l'ISO/CEI 10646 remonte à un projet de Xerox en 1986 dont le but consistait à créer une base de données croisée des caractères han. En 1988, une initiative parallèle vit le jour chez Apple ; elle s'inspirait du Thésaurus CJC du Research Libraries Group utilisé pour tenir à jour l'EACC. La fusion des bases de données de Xerox et d'Apple en 1989 a abouti au premier projet de jeu de caractères han d'Unicode. En septembre 1989, lors de la réunion du comité X3L2 chargé de la codification des jeux de caractères au sein de l'Institut de normalisation national américain, l'ANSI, le groupe de travail d'Unicode proposa que ce jeu soit inclus dans l'ISO 10646.

La principale différence entre le répertoire de caractères han d'Unicode et les tentatives antérieures résidait dans le fait que ce nouveau répertoire évinçait les différents jeux de caractères bibliographiques précédents en englobant à la fois les standards de l'industrie et les dernières normes nationales. Les critères d'unification utilisés lors de l'élaboration de ce premier répertoire de caractères han d'Unicode s'inspiraient de ceux utilisés par le JIS et d'un ensemble de principes d'identité de caractères han (*rentong yuanze*) en cours de définition par des experts chinois membres de l'Association pour un codage chinois commun (ACCC). Un des principes fondamentaux étant de conserver toutes les distinctions de caractère entre les jeux issus de l'industrie ou de normes officielles qu'ils soient en existence ou proposés.

La proposition han d'Unicode a suscité l'intérêt pour un jeu unifié han susceptible d'être incorporé à l'ISO 10646. Une réunion spéciale eut donc lieu à Pékin en octobre 1989 dont l'ordre du jour était principalement l'unification. Cette réunion marqua le début d'une coopération officielle entre le groupe de travail Unicode et l'ACCC sur les propositions respectives de chaque groupe.

En février 1990, une réunion spéciale sur l'unification han s'est tenue à Séoul. Lors de cette réunion, la délégation coréenne proposa la création d'un groupe mixte composé de représentants des pays de l'Extrême-Orient et d'autres organisations intéressées par l'étude d'un codage unifié han. Cette réunion scella la constitution d'un Groupe de recherche mixte chinois/japonais/coréen, désigné par la suite par l'abréviation GRM-CJC.

Un second projet de répertoire de caractères han Unicode fut distribué à un large public en décembre 1990 afin que sa publication coïncide avec l'annonce de la fondation du consortium Unicode. Ce projet de jeu de caractères han Unicode se distinguait du premier par son respect du principe radical-nombre de traits qui ordonne les caractères du K'ang-hsi.

Afin de confirmer de façon indépendante le bien-fondé et l'exactitude de l'unification proposée, le consortium s'assura que le projet soit minutieusement examiné par des érudits spécialistes de l'Extrême-Orient à l'Université de Toronto.

Entre-temps, la Chine annonça qu'elle était sur le point de finaliser sa propre proposition d'un jeu de caractères han, le GB 13000. Convaincus de la similarité des deux projets, tant par leur contenu que leur philosophie, le consortium Unicode et le Centre de recherche et de développement informatique (CCID) du ministère chinois de l'industrie électronique décidèrent de conjuguer leurs efforts et de présenter un projet commun. Chacun ajouta les caractères qui manquaient à l'autre proposition. On décida également d'un mécanisme pour ordonner les caractères en utilisant un système reposant sur quatre dictionnaires, ce mécanisme est décrit à la section 11.1, *Han*. Une comparaison informatique des deux bases de données de caractères bonifia considérablement les deux projets.

En vertu de l'accord de fusion entre l'ISO 10646 et le standard Unicode, le consortium Unicode adopta le répertoire de caractères han unifié sur lequel travaillait le GRM-CJC.

En juillet 1991 se tint à Tôkyô la première réunion du GRM-CJC. Le groupe se rendit compte qu'il existait un besoin patent d'unifier les caractères idéophonographiques CJC en une norme de codage cohérente. On prit donc deux décisions principales : l'utilisation du GB 13000 (fusionné au préalable avec le répertoire han d'Unicode) comme base de ce qui allait devenir le « Répertoire et classement unifiés » et la vérification des résultats de cette unification à la lumière des règles établies par le professeur Miyazawa Akira et par d'autres membres de la délégation japonaise.

L'examen officiel du GB 13000 commença aussitôt. D'autres réunions eurent lieu à Pékin et Hong Kong. Le 27 mars 1992, le GRM-CJC terminait la version 2.0 du Répertoire et classement unifiés (RCU). Ce répertoire fut publié par la suite à la fois par le consortium Unicode dans son standard Unicode, version 1.0, volume 2, et par l'ISO au sein de l'ISO/CEI 10646-1 : 1993.

En octobre 1993, le GRM-CJC devint un sous-groupe officiel du JTC1/SC2/GT2 de l'ISO/CEI et prit le nom de Groupe à rapporteur sur les idéophonogrammes (GRI). Le GRI est dorénavant officiellement chargé de publier des extensions au RCU 2.0 afin d'enrichir le répertoire codé des idéophonogrammes CJC unifiés. Le consortium Unicode participe à ce groupe en tant que membre de liaison de l'ISO.

Lors de sa deuxième réunion à Hanoï en février 1994, le GRI convenait d'inclure les idéophonogrammes vietnamiens chũ' nôm dans une version ultérieure du RCU et d'ajouter une cinquième dictionnaire de référence au mécanisme de classement.

En 1998, le GRI rendait public le premier supplément idéophonographique au RCU, le supplément A aux idéophonogrammes CJC unifiés. Ce jeu de 6.582 caractères, colligé à partir de différentes sources littéraires historiques et de normes nationales ou industrielles, fait partie de l'ISO/CEI 10646-1 : 2000 et d'Unicode 3.0. Le supplément A aux idéophonogrammes CJC unifiés représente le dernier ensemble d'idéogrammes CJC ajouté au PMB.

À l'heure actuelle (été 2000), le GRI considère l'ajout de plusieurs milliers de caractères han proposés par la Chine, Taïwan, Hong Kong, le Japon, Singapour et le Viêt-Nam au répertoire idéophonographique de l'ISO 10646.