

---

---

**Aides à apporter par les ordinateurs dans les  
travaux de terminologie — Création et  
utilisation de bases de données  
terminologiques et de corpus de textes**

iTeh STANDARD PREVIEW

(standards.iteh.ai)

*Computational aids in terminology — Creation and use of terminological  
databases and text corpora*

<https://standards.iteh.ai/catalog/standards/sist/448bde85-64be-4e58-9be9-5b6a340066f2/iso-tr-12618-1994>



## Sommaire

	Page
Avant-propos .....	iii
Introduction .....	iv
<b>1</b> Domaine d'application .....	<b>1</b>
<b>2</b> Références .....	<b>1</b>
<b>3</b> Définitions .....	<b>1</b>
<b>4</b> Types d'ensembles de données terminologiques .....	<b>2</b>
<b>5</b> Critères de création d'une base de données terminologiques .....	<b>2</b>
<b>6</b> Besoins de matériel et de logiciel .....	<b>3</b>
<b>7</b> Catégories de données terminologiques .....	<b>3</b>
<b>8</b> Structure des données .....	<b>4</b>
<b>9</b> Entrée des données .....	<b>7</b>
<b>10</b> Jeux de caractères .....	<b>8</b>
<b>11</b> Extraction des données .....	<b>8</b>
<b>12</b> Tri .....	<b>11</b>
<b>13</b> Production d'imprimés et de vocabulaires imprimés .....	<b>12</b>
<b>14</b> Protection des données .....	<b>12</b>
<b>15</b> Transfert des données .....	<b>12</b>
<b>16</b> Réaction des utilisateurs .....	<b>12</b>
<b>17</b> Maintenance et mise à jour .....	<b>12</b>
<b>18</b> Portabilité .....	<b>13</b>
<b>19</b> Transmission des données .....	<b>13</b>
<b>20</b> Création et utilisation d'un corpus de textes .....	<b>13</b>

© ISO 1994

Droits de reproduction réservés. Sauf prescription différente, aucune partie de cette publication ne peut être reproduite ni utilisée sous quelque forme que ce soit et par aucun procédé, électronique ou mécanique, y compris la photocopie et les microfilms, sans l'accord écrit de l'auteur.

Organisation internationale de normalisation  
Case postale 56 • CH-1211 Genève 20 • Suisse

Version française tirée en 1995  
Imprimé en Suisse

## Avant-propos

L'ISO (Organisation internationale de normalisation) est une fédération mondiale d'organismes nationaux de normalisation (comités membres de l'ISO). L'élaboration des Normes internationales est en général confiée aux comités techniques de l'ISO. Chaque comité membre intéressé par une étude a le droit de faire partie du comité technique créé à cet effet. Les organisations internationales, gouvernementales et non gouvernementales, en liaison avec l'ISO participent également aux travaux. L'ISO collabore étroitement avec la Commission électrotechnique internationale (CEI) en ce qui concerne la normalisation électrotechnique.

La tâche principale des comités techniques de l'ISO est d'élaborer les Normes internationales. Exceptionnellement, un comité technique peut proposer la publication d'un rapport technique de l'un des types suivants:

- type 1: lorsque, en dépit de maints efforts, l'accord requis ne peut être réalisé en faveur de la publication d'une Norme internationale;
- type 2: lorsque le sujet en question est encore en cours de développement technique ou lorsque, pour toute autre raison, la possibilité d'un accord pour la publication d'une Norme internationale peut être envisagée pour l'avenir mais pas dans l'immédiat;
- type 3: lorsqu'un comité technique a réuni des données de nature différente de celles qui sont normalement publiées comme Normes internationales (ceci pouvant comprendre des informations sur l'état de la technique, par exemple).

Les rapports techniques des types 1 et 2 font l'objet d'un nouvel examen trois ans au plus tard après leur publication afin de décider éventuellement de leur transformation en Normes internationales. Les rapports techniques du type 3 ne doivent pas nécessairement être révisés avant que les données fournies ne soient plus jugées valables ou utiles.

L'ISO/TR 12618, rapport technique du type 3, a été élaboré par le comité technique ISO/TC 37, *Terminologie (principes et coordination)*, sous-comité SC 3, *Aides à apporter par les ordinateurs dans les travaux de terminologie*.

## Introduction

Comme le présent Rapport technique se limite aux aides informatiques utilisées pour la réalisation de travaux terminologiques, il convient que l'utilisateur consulte l'ISO 704 et l'ISO 1087 pour les principes de base en terminologie.

En plus des conseils qui sont donnés dans le présent Rapport technique sur la création et l'utilisation de bases de données terminologiques, il serait bon de consulter l'ISO 6156 et l'ISO 12200 sur le format d'échange normalisé des données terminologiques et lexicographiques.

Les ordinateurs peuvent être utilisés à divers stades de la préparation et de l'utilisation d'ensembles de données terminologiques. La préparation d'ensemble de données terminologiques suppose les étapes suivantes:

- a) définition de la portée;
- b) désignation, choix et enregistrement des sources;
- c) collecte des termes, définitions, explications, exemples de textes, etc.;
- d) élaboration des systèmes de notions;
- e) établissement des relations d'équivalence entre les notions en deux langues ou plus;
- f) enregistrement de l'information terminologique, y compris l'information sur les systèmes de notions;
- g) mise à jour des données terminologiques.

Ces étapes sont énumérées dans l'ordre chronologique du processus, mais elles se chevauchent souvent et chacune peut devoir être reprise. Selon le type de projet et de ressources, les ordinateurs peuvent se révéler utiles dans bon nombre d'étapes, en particulier b), c), f) et g).

L'utilisation automatisée d'ensembles de données terminologiques comprend l'extraction de l'information terminologique stockée dans une base de données ainsi que la production d'imprimés et de dictionnaires.

Dans le présent Rapport technique, l'accent est mis sur la création et la maintenance d'une base de données terminologiques [c'est-à-dire les étapes f) et g)]. Une brève introduction à la création et à l'utilisation d'un corpus de textes pouvant servir à l'étape c) est donnée à l'article 20; il ne faut pas oublier que la création d'un corpus de textes précède la création d'une base de données terminologiques. L'information sur l'étape b), toutefois, dépasse le domaine d'application du présent Rapport technique.

# Aides à apporter par les ordinateurs dans les travaux de terminologie — Création et utilisation de bases de données terminologiques et de corpus de textes

## 1 Domaine d'application

Le présent Rapport technique fournit des conseils sur les principes de base et les méthodes qu'il faut respecter pour avoir recours à l'informatique dans la préparation et l'utilisation d'ensembles de données terminologiques. Il vise particulièrement la création et l'utilisation de bases de données terminologiques et de corpus de textes.

## 2 Références

Les normes suivantes contiennent des dispositions qui, par suite de la référence qui en est faite, constituent des dispositions valables pour le présent Rapport technique. Au moment de la publication, les éditions indiquées étaient en vigueur. Toute norme est sujette à révision et les parties prenantes des accords fondés sur le présent Rapport technique sont invitées à rechercher la possibilité d'appliquer les éditions les plus récentes des normes indiquées ci-après. Les membres de la CEI et de l'ISO possèdent le registre des Normes internationales en vigueur à un moment donné.

ISO 704:1987, *Principes et méthodes de la terminologie*.

ISO 860:1994, *Travaux de terminologie — Harmonisation internationale des notions et des termes*.

ISO 1087:1990, *Terminologie — Vocabulaire*.

ISO 1087-2:—<sup>1)</sup>, *Travaux de terminologie — Vocabulaire — Partie 2: Aides à apporter par les ordinateurs dans les travaux de terminologie*.

ISO/CEI 2382-1:1993, *Technologies de l'information — Vocabulaire — Partie 1: Termes fondamentaux*.

ISO 2382-4:1987, *Systèmes de traitement de l'information — Vocabulaire — Partie 4: Organisation des données*.

ISO 6156:1987, *Format d'échange sur bande magnétique des données terminologiques et/ou lexicographiques (MATER)*.

ISO/TR 8393:1985, *Documentation — Règles de classement bibliographique de l'ISO (Règles standards internationales de classement bibliographique) — Concrétisation des principes de classement bibliographique par des règles modèles*.

ISO 8777:1993, *Information et documentation — Commandes pour les systèmes interactifs de recherche d'information*.

ISO 8879:1986 [+ Amd 1:1988], *Traitement de l'information — Systèmes bureautiques — Langage normalisé de balisage généralisé (SGML)*.

ISO/CEI 9075:1992, *Technologies de l'information — Langages de base de données — SQL*.

ISO 10241:1992, *Normes terminologiques internationales — Élaboration et présentation*.

ISO 12200:—<sup>1)</sup>, *Aides à apporter par les ordinateurs dans les travaux de terminologie — Format d'échange terminologique (TIF) — Une application SGML*.

## 3 Définitions

Pour les besoins du présent Rapport technique, les définitions suivantes s'appliquent.

NOTE 1 La plupart de ces définitions seront intégrées à l'ISO 1087-2 et elles sont encore provisoires.

### 3.1 banque de données

ensemble de bases de données comprenant le cadre organisationnel qui les gère

NOTE 2 Voir aussi ISO/CEI 2382-1:1993.

1) À publier.

**3.2 base de données**

ensemble de données organisé selon une structure conceptuelle

NOTE 3 Adaptée de l'ISO/CEI 2382-1:1993.

**3.3 catégorie de données**

type d'éléments de données  
instruction pour interpréter une zone de données déterminée

**3.4 élément de données**

plus petite unité identifiable du contenu dans un enregistrement donné

**3.5 zone de données**

portion de longueur variable ou fixe d'un enregistrement renfermant un élément de données particulier

NOTE 4 Adaptée de l'ISO 6156:1987.

**3.6 enregistrement**

ensemble d'éléments de données considéré comme un tout [ISO 2382-4:1987]

**3.7 terminographie**

consignation, traitement et présentation des données terminologiques

NOTE 5 Adaptée de l'ISO 1087:1990.

**3.8 banque de termes**

banque de données terminologiques  
banque de données renfermant des données terminologiques

**3.9 base de données terminologiques**

base de données renfermant des données terminologiques

**3.10 ensemble de données terminologiques**

ensemble de données contenant de l'information sur des notions de domaines précis

**3.11 entrée terminologique**

partie d'un ensemble de données terminologiques qui renferme des données terminologiques reliées à une seule notion

NOTE 6 Voir aussi ISO 1087:1990, paragraphe 6.2.2.2.

**3.12 corpus de textes**

corpus  
ensemble systématique de données exploitables par machine, de textes ou de parties de texte, préparés, codés et stockés selon des règles prédéfinies

NOTE 7 Un corpus de textes peut être limité selon les aspects des domaines, volumes ou temps; par exemple des textes mathématiques, ou certains périodiques publiés depuis 1986. Il est utilisé en tant que source pour des travaux ultérieurs d'analyse linguistique ou de terminologie.

NOTE 8 Voir aussi ISO 1087:1990, paragraphe 6.1.2.2.

**4 Types d'ensembles de données terminologiques**

Les critères suivants influent sur la manipulation des ensembles de données terminologiques et sur leur accès:

- **volume:** nombre d'entrées, domaines, langues, types de données;
- **matériel:** micro-ordinateur, mini-ordinateur, ordinateur central; disque dur, disquette, CD-ROM; système autonome ou réseau;
- **logiciel:** système de gestion de base de données, système de recherche documentaire, système d'édition de dictionnaire; conception standard ou personnalisée;
- **propriétaire/utilisateur:** organisation internationale, institution nationale, société, particulier; accès libre ou restreint;
- **applications:** extraction en ligne ou hors ligne des termes pour la traduction assistée par ordinateur, imprimés (par exemple contenant toutes les entrées d'un domaine comme matériel de base à l'intention d'un groupe de travail), production de vocabulaires imprimés (composition par ordinateur); utilisation dans des systèmes experts ou dans des systèmes de traduction machine.

D'autres types d'ensembles de données peuvent être intégrés à des ensembles de données terminologiques, par exemple:

- bases de données de textes intégraux (voir aussi article 20);
- bases de données graphiques;
- bases de données numériques;
- bases de données bibliographiques.

**5 Critères de création d'une base de données terminologiques**

La création d'une base de données terminologiques peut être utile si un ou plusieurs des critères ci-dessous sont remplis:

- a) Il faut uniformiser la terminologie monolingue, bilingue ou multilingue au niveau international ou national ou au niveau de l'entreprise.
- b) Il existe un besoin permanent de mise à jour ou de révision de volumes importants de données.
- c) Il existe un besoin de consultation de données terminologiques au moyen de critères différents ou de combinaisons de critères (par exemple, par terme dans une langue, par sujet ou par source).
- d) Il existe un besoin de présentation des données dans des formats différents selon les spécifications de l'utilisateur (par exemple, vocabulaires

spéciaux classés alphabétiquement ou systématiquement comme sous-ensembles pour traduction machine ou assistée par ordinateur).

- e) Le nombre d'utilisateurs éventuels nécessitant un accès rapide aux données est suffisamment élevé pour justifier l'investissement en matériel, logiciel et ressources humaines (formation, programmation, maintenance, etc.).
- f) Les ressources humaines nécessaires sont disponibles pour la formation du personnel et pour la création et la maintenance de la base de données, tout comme le sont les ressources financières pour l'acquisition du matériel et du logiciel.

## 6 Besoins de matériel et de logiciel

Le volume de l'ensemble de données terminologiques et le nombre d'utilisateurs éventuels déterminent s'il faut un micro-ordinateur, un mini-ordinateur ou un ordinateur central.

On peut se servir de divers types de logiciels pour enregistrer et utiliser des ensembles de données terminologiques, par exemple des systèmes de traitement de texte, des systèmes d'édition de dictionnaire, des systèmes de gestion de base de données et des systèmes de recherche documentaire. Les systèmes de gestion de base de données — et dans une certaine mesure les systèmes de recherche documentaire — sont les systèmes les plus souples pour la manipulation des données. Par conséquent, le présent Rapport technique met l'accent sur la création et l'utilisation d'une base de données terminologiques à l'aide d'un système de gestion de base de données ou d'un système de recherche documentaire. Dans le texte qui suit, l'expression «système de base de données» vaut tant pour les systèmes de gestion de base de données que pour les systèmes de recherche documentaire.

Beaucoup de systèmes de base de données sont disponibles pour différents systèmes d'exploitation, en versions monutilisateur ou multiutilisateur. Certains systèmes tournent sur micro-ordinateur, mini-ordinateur et ordinateur central. Idéalement, il devrait être possible de faire passer le système à une puissance supérieure (de la version micro-ordinateur à la version miniordinateur ou de la version monutilisateur à la version multiutilisateur, voir article 18).

Certains systèmes tournent sur des micro-ordinateurs dotés d'une mémoire interne très limitée, mais il est souvent recommandable d'investir dans de l'espace mémoire vive supplémentaire. Si le système dialogue avec d'autres programmes et fait ainsi partie intégrante d'un système plus puissant, il faut encore davantage d'espace mémoire.

La plupart des systèmes de base de données ont besoin d'espace mémoire supplémentaire pour la gestion des données (marques internes, index, etc.), par-

fois jusqu'à 10 fois l'espace requis pour le texte «brut» des entrées terminologiques. Par exemple, si le volume moyen d'une entrée est de 1 000 caractères (octets), 1 000 entrées peuvent occuper jusqu'à 10 Mo d'espace mémoire, bien que beaucoup de systèmes aient des fonctions qui réduisent l'espace occupé par la base de données. Il devrait y avoir des fonctions de sauvegarde, sur disque dur ou sur disquette, ou sur une unité à bande magnétique.

Une base de données terminologiques peut être accessible sur support optique, par exemple sur disque compact CD-ROM, qui peut contenir de grandes quantités de données. Il faut du matériel spécial pour consulter une base de données sur support optique.

## 7 Catégories de données terminologiques

La structure et les types de données des entrées terminologiques doivent être clairement décrits. Ces descriptions sont nécessaires à la manipulation des données.

Il convient que les types de données soient définis et délimités indépendamment. Ainsi, chaque élément de données peut être attribué sans ambiguïté à un seul type. La «classification des domaines» est un exemple de type de données. «CDU 621» ou «CDU 347» pourraient donc être des éléments de données pertinents.

Voici d'autres exemples de types de données:

- terme;
- information grammaticale;
- définition;
- contexte;
- collocation;
- relation entre notions;
- sources.

Des ensembles de données différents nécessitent des combinaisons différentes de types de données. Une base de données qui sert à produire des dictionnaires imprimés contient des types de données différents de ceux d'une base servant à la recherche de termes individuels. Des groupes d'utilisateurs différents (par exemple, des étudiants, des traducteurs, des experts) ont besoin d'informations différentes.

Bien souvent, une base de données terminologiques est multifonctionnelle. Toutefois, il n'est pas toujours possible de prévoir tous les besoins à l'étape de la planification. Il est donc conseillé de définir une structure de base de données qui soit la plus souple possible afin de pouvoir ajouter de nouveaux types de données à n'importe quelle étape.

## 8 Structure des données

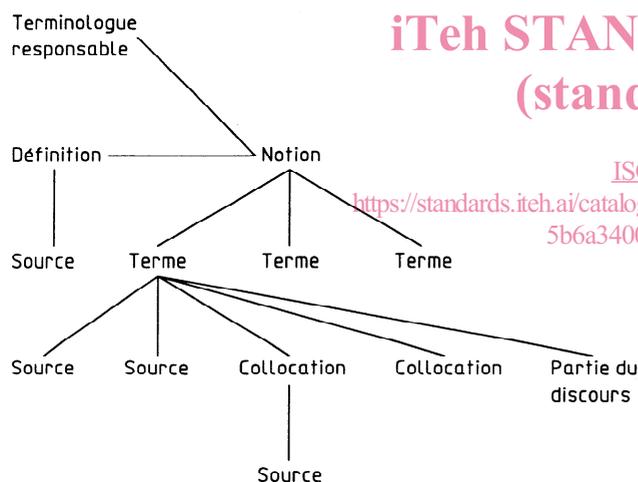
### 8.1 Structure des données terminologiques

Afin de pouvoir décrire la structure des entrées dans un ensemble de données terminologiques, il faut avoir de l'information sur les relations entre les éléments de données.

Chaque élément de données est relié à la notion dans son ensemble ou à tout autre élément de donnée, généralement un terme. Les éléments de données peuvent être facultatifs ou obligatoires et on doit pouvoir les répéter ou non.

Le format interne des données diffère normalement du format externe, c'est-à-dire le format présenté à l'utilisateur.

Certains éléments de données comme la définition, le terminologue responsable, etc., sont habituellement reliés à la notion. La partie du discours, les collocations, etc., sont reliés aux termes. La source peut être reliée aux définitions, aux termes, aux collocations, etc. Ces relations sont illustrées à la figure 1.



**Figure 1 — Éléments de données dans une entrée unilingue avec trois termes synonymes**

Les termes de la figure 1 sont des synonymes qui peuvent être accompagnés d'information sur les restrictions d'usage stylistique ou régional. Il faut souvent ajouter certains éléments à chaque terme (par exemple sources, collocations, notes, etc.). Les termes illustrés à la figure 1 pourraient aussi être organisés en un seul terme privilégié (dans la zone TERME) et en deux termes admis (dans une zone SYNONYME). Cette méthode s'applique à une base de données terminologiques pour une terminologie uniformisée.

Chaque entrée terminologique renferme de l'information sur une notion dans une ou plusieurs langues. Il peut toutefois être valable d'inclure des notions partiellement équivalentes en deux langues dans la même entrée si, en dépit des différences, il est rai-

sonnable d'utiliser les termes comme traduction l'un de l'autre. Dans ce cas, les différences entre les deux notions doivent être clairement indiquées dans une note sur l'équivalence.

Dans des domaines où il n'y a pas de différences importantes d'équivalence — ce qui est souvent le cas dans les domaines techniques — une entrée peut contenir de l'information dans plus de deux langues. Dans les domaines comme le droit, les sciences sociales, l'éducation, etc., les différences d'équivalence rendent souvent l'approche multilingue impossible. Dans ce cas, il vaut mieux que l'information de chaque entrée se rapporte à une seule paire de langues. Idéalement, il devrait être possible de stocker une terminologie unilingue, bilingue et multilingue dans la même base de données.

Une base de données peut se composer de paires de langues, une des langues étant toujours la même, par exemple l'anglais dans une banque de termes anglais. Si l'on veut permettre des recherches dans deux autres langues, par exemple le français et l'allemand, il est conseillé de prévoir des restrictions ou des avertissements automatiques signalant que la relation d'équivalence n'est établie qu'entre une langue donnée et l'anglais. Par exemple, si une notion anglaise est reliée à une notion allemande et à une notion française, et s'il y a une équivalence partielle dans les deux paires de langues (anglais—allemand et anglais—français), l'utilisateur doit être informé que la relation d'équivalence entre l'allemand et le français n'a pas été vérifiée.

Idéalement, la terminologie d'un domaine doit être établie parallèlement en deux langues ou plus. Les systèmes de notions sont établis et les notions sont définies indépendamment pour toutes les langues. Les relations d'équivalence entre les langues sont ensuite établies en comparant les définitions et les systèmes de notions (voir ISO 860). Toutes les catégories d'information, définitions, contextes, sources, etc., peuvent être fournies pour chaque notion dans toutes les langues. Par conséquent, aucune langue n'est considérée comme langue-source ou une langue-cible dans la base de données.

Lorsque la base de données est utilisée pour l'interrogation interactive de termes ou pour la production d'un vocabulaire imprimé, chaque langue peut être choisie comme langue-source ou langue-cible, et il est possible de choisir différents sous-ensembles d'information selon le groupe d'utilisateurs et l'objet du dictionnaire (voir article 11).

Lorsqu'une notion dans une langue n'a pas d'équivalent dans une autre langue, une traduction peut être suggérée, mais cette dernière ne doit jamais figurer comme terme de la langue-source dans un dictionnaire imprimé. Par conséquent, ces traductions proposées doivent être signalées comme telles dans la base de données.

### 8.2 Établissement d'une base de données

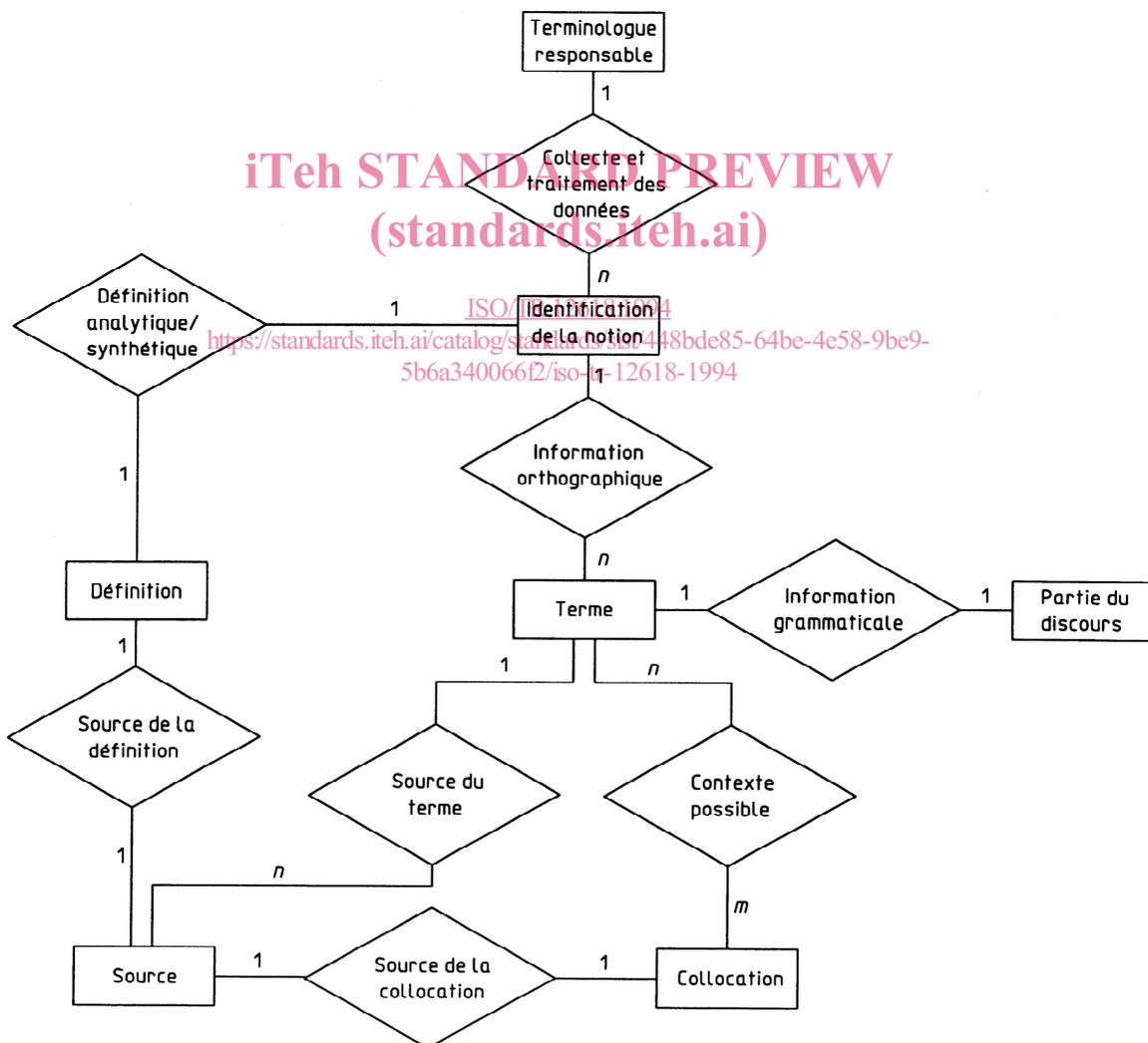
Pour établir une base de données terminologiques, il faut une description formelle de la structure des données. À cette fin, on peut utiliser divers types de diagrammes. L'un d'eux est le diagramme entité-relation qui sert à décrire la structure des données dans une base de données hiérarchique ou relationnelle ou encore dans une base de données en réseau. La structure arborescente simplifiée d'une entrée terminologique de la figure 1 peut être représentée dans un diagramme entité-relation comme l'illustre la figure 2 dans laquelle on peut observer les types de relations suivants:

- relations un-à-un (1:1), par exemple entre terme et partie du discours;
- relations un-à-plusieurs (1:n), par exemple entre notion et terme;

- relations plusieurs-à-plusieurs (n:m), par exemple entre terme et collocation.

La structure d'une entrée terminologique est mise en place de façons différentes dans divers systèmes. La figure 3 illustre une des façons possibles, dans un système de base de données relationnelle, de mettre en place la structure de données (illustrée à la figure 2).

Dans une base de données relationnelle, l'entrée terminologique est divisée en plusieurs enregistrements dans diverses tables interreliées. Par exemple, la relation entre une notion et un ou plusieurs synonymes est établie au moyen du numéro d'identification et de la LANGue. Pendant l'extraction, les éléments de données reliés à une même entrée terminologique sont regroupés et présentés comme une seule unité.



**Explication des symboles**

□ Type d'entité    ◇ Type de relation

**Figure 2 — Exemple de diagramme entité-relation d'une base de données terminologiques**

**concept**

ID	LANG	RESP
60201	da	HPL
60201	fr	HPL
60201	es	MMJ
60204	da	HPL
60204	fr	HPL
60204	es	MMJ

**def**

ID	LANG	DEF
60201	da	Juridisk er forsikring en aftale, hvor den ene part, forsikringsgiveren, forpligter sig til at udbetale en erstatning til den anden part, forsikringstageren, såfremt en af aftelen omfattet begivenhed indtræder. Som modydelse betaler forsikringstageren en præmie.
60201	fr	Une opération par laquelle une partie, l'assuré, se fait promettre, moyennant une rémunération, la prime, pour lui ou pour un tiers, en cas de réalisation d'un risque, une prestation par une autre partie, l'assureur, qui, prenant en charge un ensemble de risques, les compense conformément aux lois de la statistique.
60204	da	Ved livsforsikring forstås dels en forsikring, hvor forsikringssummen udbetales ved eller en bestemt tid efter en persons død, og dels en forsikring, hvor summen udbetales i levende live, fx. ved opnåelse af en bestemt alder eller ved indgåelse af ægteskab.
60204	fr	Les assurances sur la vie sont destinées à garantir, soit le risque de mort de la personne assurée (assurance en cas de décès), soit le risque de sa survie à une époque déterminée (assurance en cas de vie).
60204	es	El seguro sobre la vida comprenderá todas las combinaciones que pueden hacerse, pactando entregas de primas o entrega de capital a cambio de disfrute de renta vitalicia o hasta cierta edad, o percibo de capitales al fallecimiento de persona cierta.

**defref**

ID	LANG	REF
60201	da	Bac p 9
60201	fr	BES p 2
60204	da	PL p 383
60204	fr	Bes p 32
60204	es	CC 416

**term**

ID	LANG	TNO	TERM
60201	da	1	forsikring
60201	fr	1	assurance
60201	es	1	seguro
60204	da	1	livsforsikring
60204	fr	1	assurance sur la vie
60204	fr	2	assurance-vie
60204	es	1	seguro sobre la vida
60204	es	2	seguro de vida

**termref**

ID	LANG	TNO	REF
60201	da	1	Bac p 9
60201	fr	1	Bes p 2
60201	es	1	MMJ
60204	da	1	PL p 383
60204	fr	1	Bes p 32
60204	fr	2	Vey II p 180
60204	es	1	CCE p 416
60204	es	2	MMJ

**coll**

ID	LANG	TNO	CNO	COLL
60201	da	1	1	tegne forsikring
60201	da	1	2	forsikringen dækker tab
60201	fr	1	1	contracter une assurance
60201	fr	1	2	conclure une assurance

**collref**

ID	LANG	TNO	CNO	REF
60201	da	1	1	Pl p 126
60201	da	1	2	Bac p 11
60201	fr	1	1	Bes p 75
60201	fr	1	2	Bes p 75

Figure 3 — Tables d'une base de données relationnelle avec exemple de données

Dans d'autres systèmes, par exemple un système de recherche documentaire, tous les éléments de données d'une entrée terminologique sont stockés dans un seul enregistrement. Quel que soit le type de système utilisé, il est très important pour la recherche interactive, la production de vocabulaires et l'échange de données que chaque élément de donnée et ses relations avec les autres éléments puissent être déterminés séparément. Si cette exigence n'est pas respectée, il n'est pas possible, par exemple, de spécifier des profils de recherche et de représentation propres à des groupes d'utilisateurs (voir 11.9).

### 8.3 Modification de la structure des données

Bien qu'il faille une structure d'entrée déterminée avant de pouvoir constituer une banque de termes, on doit pouvoir apporter des changements à la structure des données en tout temps.

Par exemple, on devrait pouvoir:

- ajouter une zone;
- réorganiser les structures hiérarchiques;
- changer l'ordre des zones;
- subdiviser ou fusionner des zones;
- changer les types de données des zones (par exemple, entier, caractère, date).

#### EXEMPLE

Dans la première version d'une base de données terminologiques, des termes synonymes sont classifiés comme un terme (zone TERME) avec un ou plusieurs synonymes (zone SYNONYME). Par la suite, on décide de classier tous les termes synonymes comme termes et de supprimer la zone SYNONYME (voir 8.1).

Idéalement, les noms de zones devraient être des abréviations mnémoriques comme TERM, DÉF, RÉF, etc.

### 8.4 Exigences quantitatives

Quelques systèmes de base de données ont des restrictions quantitatives qui sont inacceptables pour le travail terminologique, dans lequel les conditions suivantes doivent être respectées:

- aucune limite au nombre d'entrées terminologiques dans la base de données (en pratique, un maximum d'environ un million d'entrées devrait suffire);

- aucune limite au nombre de zones (éléments de données) par entrée;
- aucune limite au nombre de caractères par zone (longueur de zone).

Un système de base de données dans lequel la longueur des zones est fixe ne conviendrait pas parce que les données terminologiques sont souvent de longueur variable et peuvent renfermer des éléments de données facultatifs.

## 9 Entrée des données

Les données peuvent être introduites par entrée interactive des données ou par entrée des données par lots, ou par une combinaison de ces entrées.

### 9.1 Entrée interactive des données

La plupart des systèmes de base de données permettent l'entrée directe des données ainsi que la mise à jour et les corrections en mode interactif. Cela signifie en général qu'on peut immédiatement avoir accès aux données corrigées. Cette forme d'entrée des données n'est pratique que lorsque le volume de données à introduire est limité. Autrement, la mise à jour a lieu à intervalles réguliers.

### 9.2 Entrée des données par lots

Les systèmes de base de données disposent généralement d'une option d'entrée par lots qui permet de transférer dans la base même des données créées extérieurement. Lorsque suffisamment de données ont été saisies à l'aide d'un utilitaire externe, elles sont transférées dans la base de données par lots. L'utilitaire externe servant à entrer les données peut être un programme de traitement de texte.

Les données terminologiques exploitables par machine et celles que l'on peut rendre exploitables, par lecteur optique par exemple, peuvent normalement être transférées dans une base de données terminologiques. Il faut généralement restructurer ces données au niveau du format d'enregistrement et des jeux de caractères. La nature et l'importance de la restructuration dépendent des données-sources. S'il s'agit de données de dictionnaires imprimés, il faut analyser le style des caractères et la ponctuation pour déterminer les catégories correspondantes. Dans certains cas, il faut développer des programmes analyseurs très sophistiqués et spécialisés.

### 9.3 Utilitaires de traitement de texte et de vérification

Il faut prévoir une importante ressource de traitement de texte pour l'entrée, l'édition, la modification et la correction des données.