Contribution of terminological resources in learning specialized languages

Isabelle Carrière

Observatoire de linguistique Sens-Texte (OLST)

Université de Montréal

Plan

1- Introduction
    Motivation and objectives
2- Why study specialized adjectives
    Derivatives of nouns
3- Methodology
4- DATUM
    a) Encoding of the data
    b) Using the encoded data
5- Conclusion
    Future research opportunities

Introduction – Motivation (1)

1. Adjectives are rarely described in specialized dictionaries
2. When they are present :
    a) **Mentioned in the entries of the nouns they are derived from** (e.g. *abdominal* → *abdomen*)
    b) **Described using a list of noun phrases where they are found** (e.g. *abcès abdominal, infection abdominale, bactérie abdominale, tumeur abdominale, chirurgie abdominale*)
    c) **Described by a simple reference to their nominal bases** (e.g. *abdominal*: related to the abdomen)

1-Introduction – Motivation (2)

Abdominal

2) Related to the abdomen

5) - *abcès abdominal*
- *infection abdominale*
- *bactérie abdominale*
- *tumeur abdominale*
- *chirurgie abdominale*

abcès abdominal
tumeur abdominale

Which is located in x

chirurgie abdominale

Which is performed on x

infection abdominale
bactérie abdominale

Which affects x

Introduction – Objectives

1) Describe adjectives while giving key information on specialized languages

3) Develop an encoding that results in a multi-level dictionary, which respond to the needs of a wide variety of users

Why describe adjectives (1)

- **In the past decades, we have seen a growing interest in adjectives.**
  - **Their role and place in specialized texts**
    - Bae 2006; L'Homme 2004; Maniez 2001; Zweigenbaum, Hadouche and Grabar 2003
  - **Their automatic processing**
    - Daille 2001; Grabar and Zweigenbaum 2003; Normand and Bourigault 2001

Why describe adjectives (2) – Sorting qualifying adjectives

Qualifying adjectives

1- Regular qualifiers

e.g. *grand – petit* ‘tall-short’

chronique – aigu ‘chronic-acute’

2- Derivatives of verbs

e.g. *adhérent*, derived from *adhérer*

3- Derivatives of nouns

e.g. *viral*, derived from *virus*

Why describe adjectives (3) – Derivatives of nouns

Like qualifiers: (Grevisse and Goosse 1993)

- Is optional in the sentence structure
- Modifies a noun
- Agrees in number and gender with the noun it modifies

Unlike qualifiers: (Monceaux 1997; Daille 2001; Goes 1999)

- Non-predicative position
- Incompatibility with adverbial modification of degree
- Semantic relation with a base noun, which often shares formal similiarities with the adjective
- Substitution of a paraphrase containing the base noun for the adjective

Why describe adjectives (4) – Derivatives of nouns

Examples:

- Non-predicative position

*Le patient est cardiaque → The noun *patient* is understood, although omitted before *cardiac*

- Adverbial modification of degree

*L'infection est très bactérienne → an infection is bactérial or not.

- Semantic relation with a noun

Souche vaccinal → strain used in a **vaccine**

Prévention vaccinal → preventive approch using **vaccination**

- Substitution of a paraphrase

Maladie infectieuse → maladie qui est causée par une infection
'disease caused by an infection'

Why describe adjectives (5) – Special cases of derivatives

4) Derivatives sharing an attenuated relation with a noun

- Often nouns used as adjectives, but with a change in its meaning.
- Behave like derivatives of nouns.

Example: *clinique* ‘clinical/clinic’

1) *clinique n.*: Private facility destined to provide outpatients care. (Our translation of Garnier et Delamare 2002 : s.v.)

2) *clinique adj.*: *signe clinique ‘clinical sign’*
~ which is observed in a patient
* ~ which is observed in a clinic

Methodology (1) - Corpora

Criteria	In our study
Language	French
Specialization	1) Medical domain 2) Experts
Occurrences	200 000 occurrences
Sub-domain	Infectious disease
Type	1) Articles 2) On-line medical textbooks 3) Technical charts
Source	1) Texts written in French

Methodology (2) – Extraction

1) extract terms (TermoStat) to identify the adjectives specific to our corpora
2) Eliminate noise
3) Select the derivatives of nouns

Methodology (3) – Analysis

1) List the head nouns of the noun phrases (NomT)
2) Analyze the meaning of the adjective in its context
3) Describe the acception (sense) observed according to the head noun and the context

Steps completed using Syntex

Methodology (4) – Description

Description based on the meaning of derivatives of nouns in NA noun phrase

Noun + **A**djective = Head noun + semantic relation + Base noun

Syntactic patterns	**Examples**
HNoun + **preposition** + BNoun	HNoun des bactéries ‘of the bacteria’ e.g. *destruction bactérienne* *‘bacterial destruction’*
HNoun + **prop. with a conjugated verb** + BNoun	HNoun qui est causé par un virus ‘which is caused by a virus’ e.g. *infection virale* ‘viral infection’

HNoun: Head noun; BNoun: Base noun

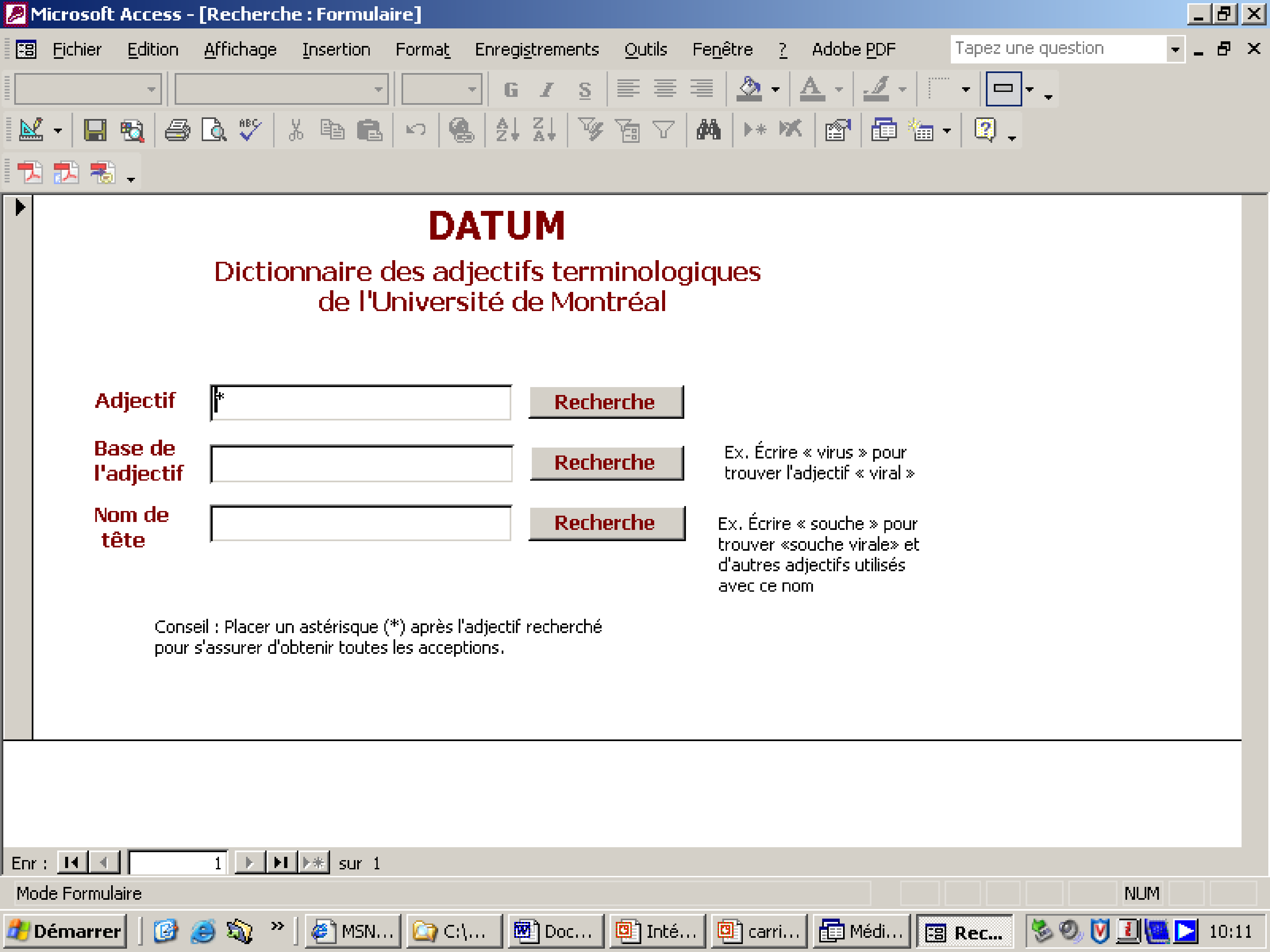
Monceaux 1997

Encoding of the data (1)

DATUM - Dictionnaire des adjectifs terminologiques de l’Université de Montréal

Terminological database that presents the various acceptions of specialized adjectives, while listing the nouns they are combined with and giving contexts in which they have been found.

Encoding of the data (2) Query form


Microsoft Access - [Recherche : Formulaire]
Fichier
Edition
Affichage
Insertion
Format
Enregistrements
Outils
Fenêtre
?
Adobe PDF
Tapez une question
Enr :
1
sur 1
Mode Formulaire
NUM
Démarrer
MSN...
C:\...
Doc...
Inté...
carri...
Médi...
Rec...
10:11

DATUM

Dictionnaire des adjectifs terminologiques de l'Université de Montréal

Adjectif * **Recherche**

Base de l'adjectif **Recherche**

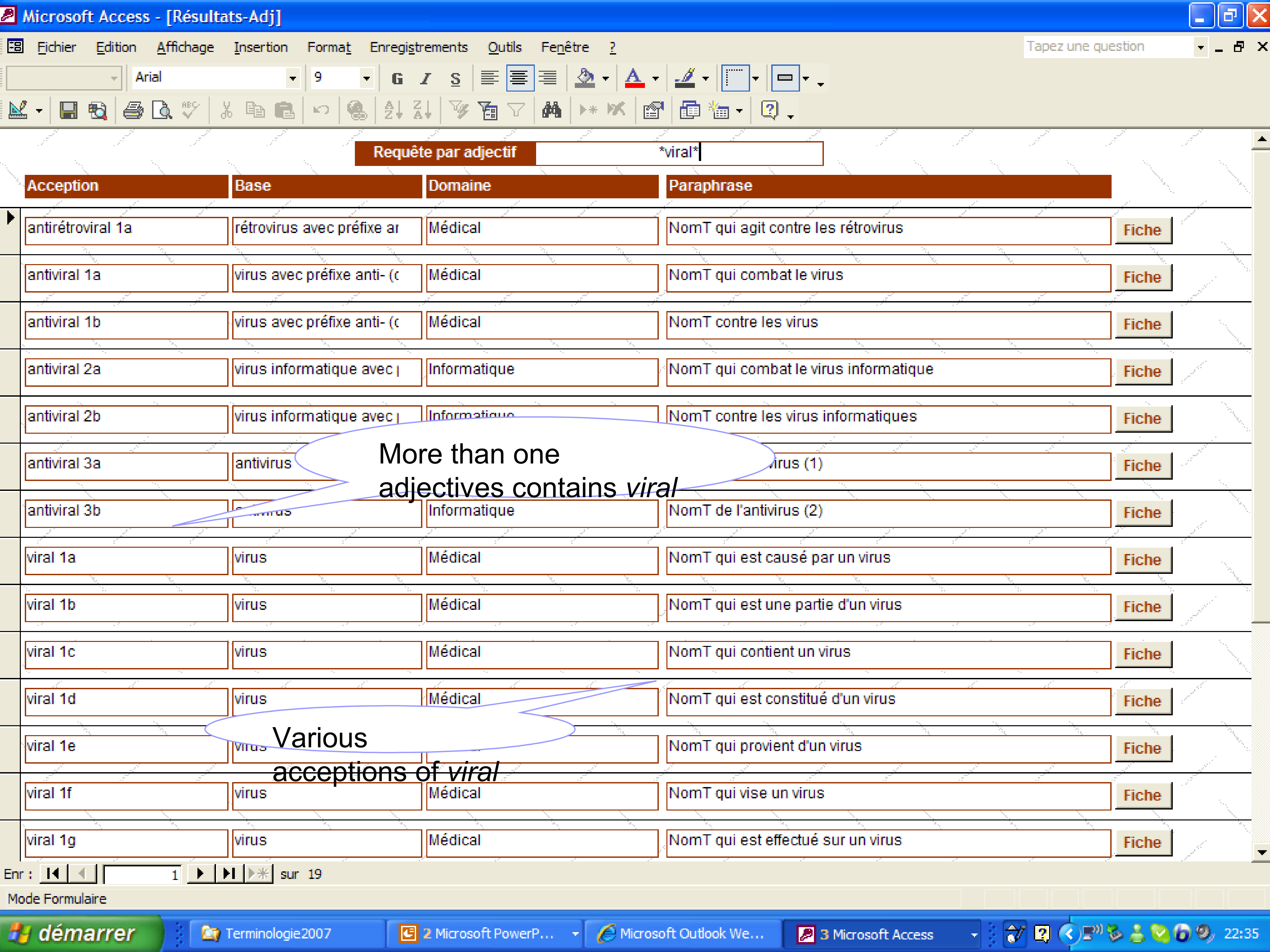
Ex. Écrire « virus » pour trouver l'adjectif « viral »

Nom de tête **Recherche**

Ex. Écrire « souche » pour trouver «souche virale» et d'autres adjectifs utilisés avec ce nom

Conseil : Placer un astérisque (*) après l'adjectif recherché pour s'assurer d'obtenir toutes les acceptions.

Encoding of the data (3) Results – Query by adjective


Microsoft Access - [Résultats-Adj]
Fichier Edition Affichage Insertion Format Enregistrements Outils Fenêtre ?
Tapez une question
Arial
9
Enr : 1 sur 19
Mode Formulaire
démarrer
Terminologie2007
2 Microsoft PowerP...
Microsoft Outlook We...
3 Microsoft Access
22:35

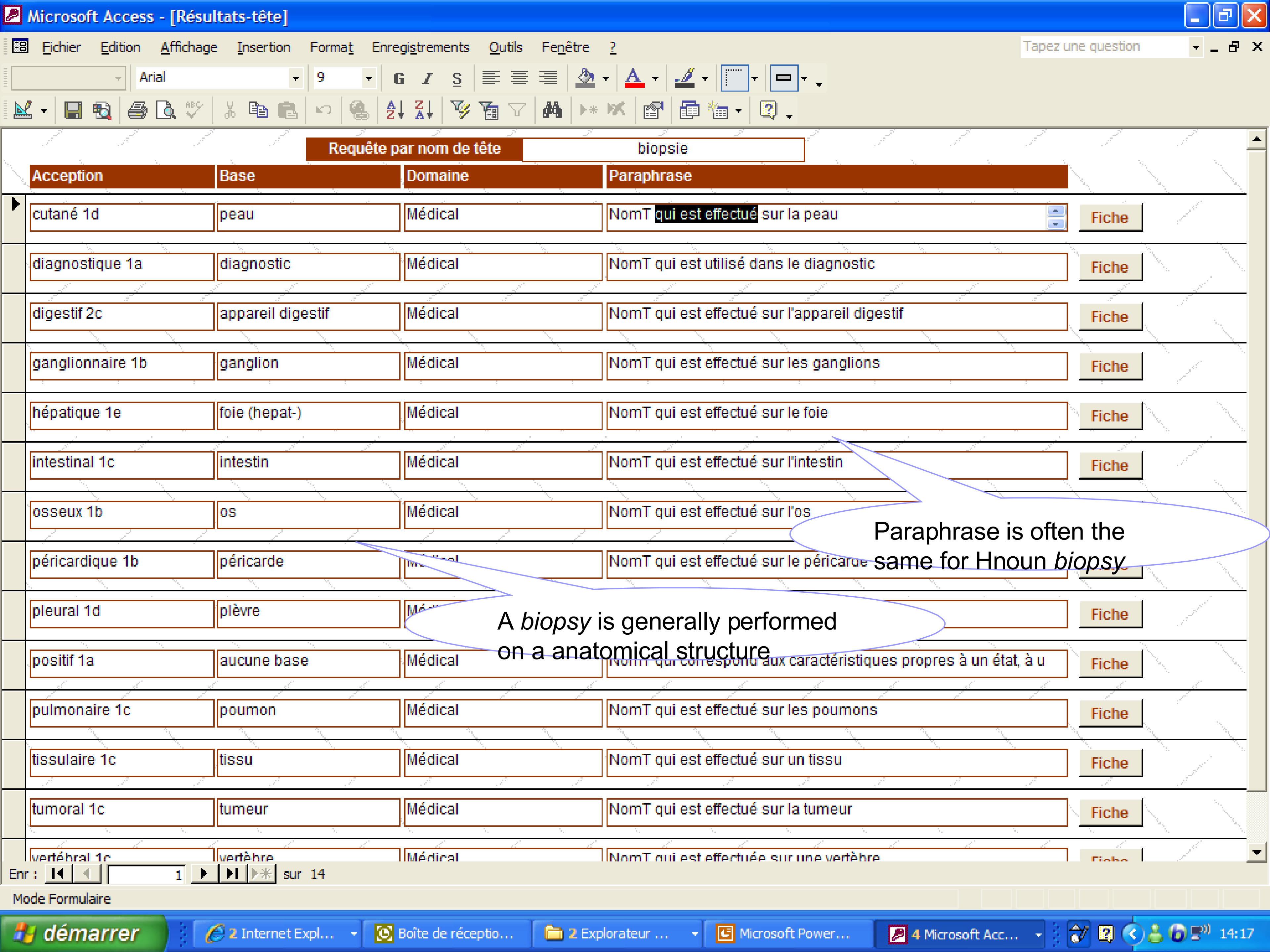
Requête par adjectif *viral*

Acception	Base	Domaine	Paraphrase	
antirétroviral 1a	rétrovirus avec préfixe ar	Médical	NomT qui agit contre les rétrovirus	Fiche
antiviral 1a	virus avec préfixe anti- (c	Médical	NomT qui combat le virus	Fiche
antiviral 1b	virus avec préfixe anti- (c	Médical	NomT contre les virus	Fiche
antiviral 2a	virus informatique avec	Informatique	NomT qui combat le virus informatique	Fiche
antiviral 2b	virus informatique avec	Informatique	NomT contre les virus informatiques	Fiche
antiviral 3a	antivirus		irus (1)	Fiche
antiviral 3b	antivirus	Informatique	NomT de l'antivirus (2)	Fiche
viral 1a	virus	Médical	NomT qui est causé par un virus	Fiche
viral 1b	virus	Médical	NomT qui est une partie d'un virus	Fiche
viral 1c	virus	Médical	NomT qui contient un virus	Fiche
viral 1d	virus	Médical	NomT qui est constitué d'un virus	Fiche
viral 1e	virus		NomT qui provient d'un virus	Fiche
viral 1f	virus	Médical	NomT qui vise un virus	Fiche
viral 1g	virus	Médical	NomT qui est effectué sur un virus	Fiche

More than one adjectives contains *viral*

Various acceptions of *viral*

Encoding of the data (4)
Results – Query by Head Noun


Microsoft Access - [Résultats-tête]
Fichier
Edition
Affichage
Insertion
Format
Enregistrements
Outils
Fenêtre
?
Tapez une question
Arial
9
Fiche
Enr :
1
sur 14
Mode Formulaire
démarrer
2 Internet Expl...
Boîte de réceptio...
2 Explorateur ...
Microsoft Power...
4 Microsoft Acc...
14:17

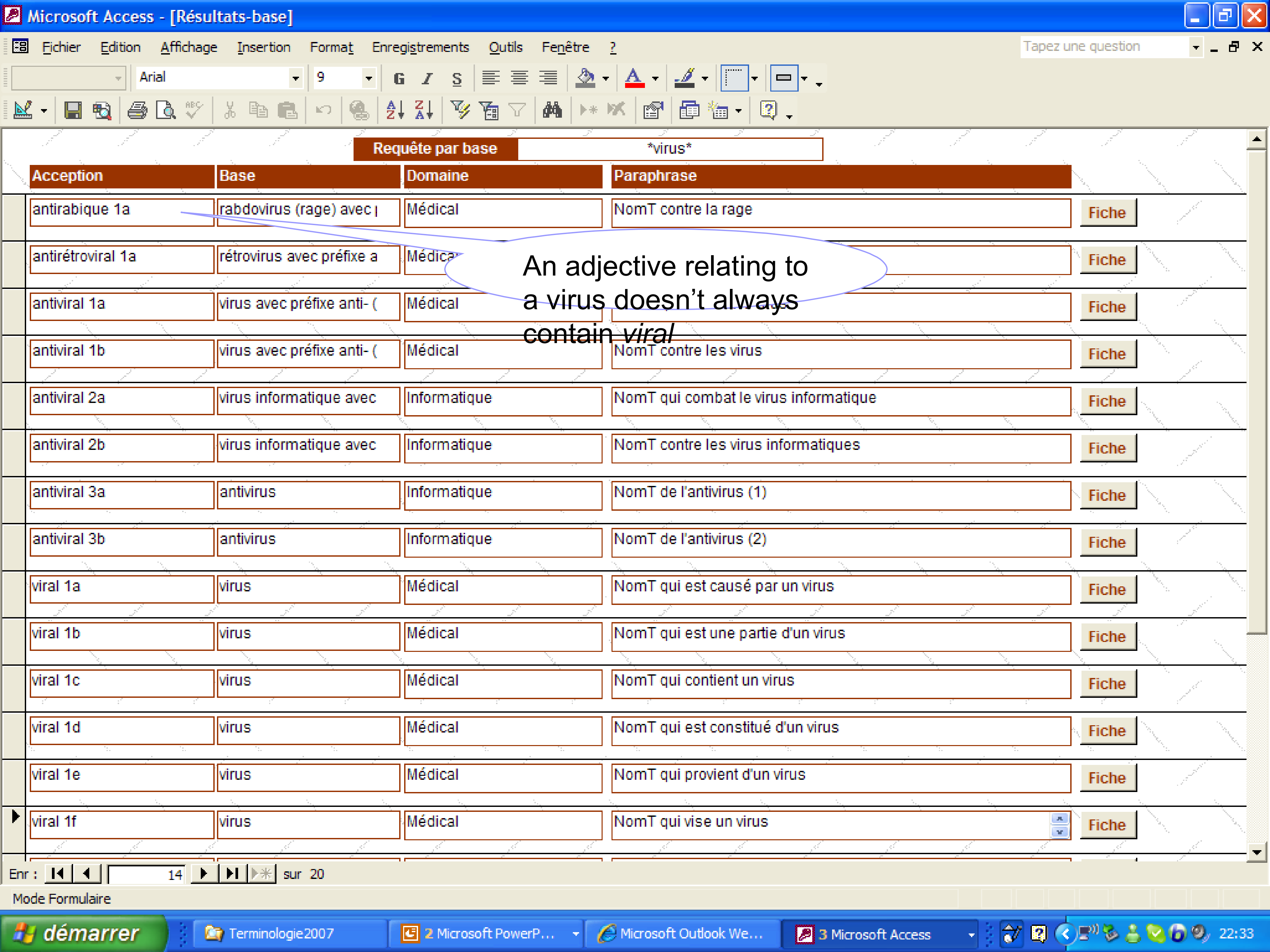
Requête par nom de tête biopsie

Acception	Base	Domaine	Paraphrase
cutané 1d	peau	Médical	NomT qui est effectué sur la peau
diagnostique 1a	diagnostic	Médical	NomT qui est utilisé dans le diagnostic
digestif 2c	appareil digestif	Médical	NomT qui est effectué sur l'appareil digestif
ganglionnaire 1b	ganglion	Médical	NomT qui est effectué sur les ganglions
hépatique 1e	foie (hepat-)	Médical	NomT qui est effectué sur le foie
intestinal 1c	intestin	Médical	NomT qui est effectué sur l'intestin
osseux 1b	os	Médical	NomT qui est effectué sur l'os
péricardique 1b	péricarde	Médical	NomT qui est effectué sur le péricarde
pleural 1d	plèvre	Mé	
positif 1a	aucune base	Médical	NomT qui correspond aux caractéristiques propres à un état, à u
pulmonaire 1c	poumon	Médical	NomT qui est effectué sur les poumons
tissulaire 1c	tissu	Médical	NomT qui est effectué sur un tissu
tumoral 1c	tumeur	Médical	NomT qui est effectué sur la tumeur
vertébral 1c	vertèbre	Médical	NomT qui est effectuée sur une vertèbre

Paraphrase is often the same for Hnoun *biopsy*

A *biopsy* is generally performed on a anatomical structure

Encoding of the data (5)
Results – Query by Base noun


Microsoft Access - [Résultats-base]
Fichier
Edition
Affichage
Insertion
Format
Enregistrements
Outils
Fenêtre
?
Tapez une question
Arial
9
démarrer
Terminologie2007
2 Microsoft PowerP...
Microsoft Outlook We...
3 Microsoft Access
22:33

An adjective relating to a virus doesn't always contain *viral*

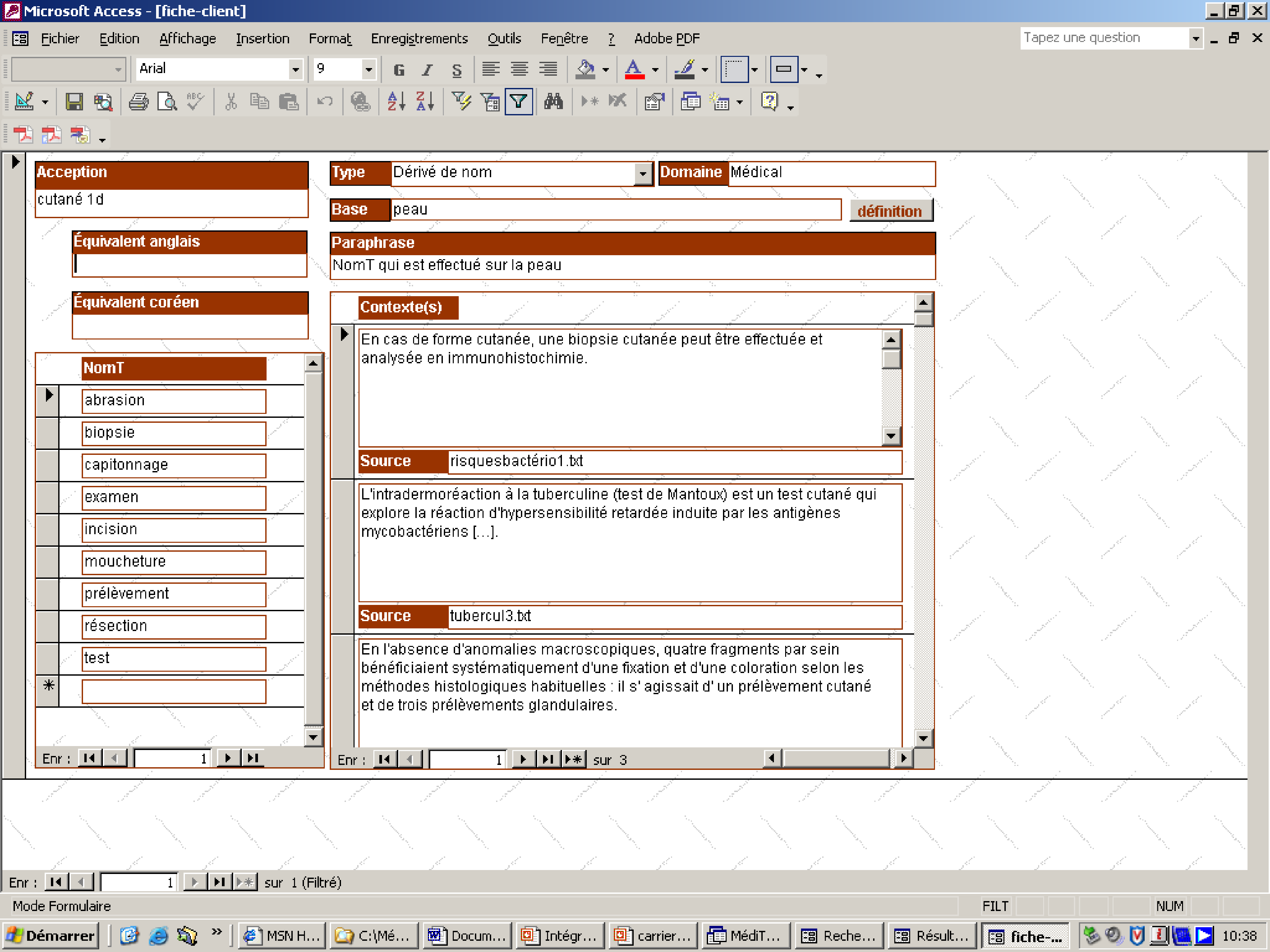
Requête par base *virus*

Acception	Base	Domaine	Paraphrase	
antirabique 1a	rabdovirus (rage) avec	Médical	NomT contre la rage	Fiche
antirétroviral 1a	rétrovirus avec préfixe a	Médica		Fiche
antiviral 1a	virus avec préfixe anti- (	Médical		Fiche
antiviral 1b	virus avec préfixe anti- (	Médical	NomT contre les virus	Fiche
antiviral 2a	virus informatique avec	Informatique	NomT qui combat le virus informatique	Fiche
antiviral 2b	virus informatique avec	Informatique	NomT contre les virus informatiques	Fiche
antiviral 3a	antivirus	Informatique	NomT de l'antivirus (1)	Fiche
antiviral 3b	antivirus	Informatique	NomT de l'antivirus (2)	Fiche
viral 1a	virus	Médical	NomT qui est causé par un virus	Fiche
viral 1b	virus	Médical	NomT qui est une partie d'un virus	Fiche
viral 1c	virus	Médical	NomT qui contient un virus	Fiche
viral 1d	virus	Médical	NomT qui est constitué d'un virus	Fiche
viral 1e	virus	Médical	NomT qui provient d'un virus	Fiche
viral 1f	virus	Médical	NomT qui vise un virus	Fiche

Enr : 14 sur 20

Mode Formulaire

Encoding of the data (6) Term record


Microsoft Access - [fiche-client]
Fichier
Edition
Affichage
Insertion
Format
Enregistrements
Outils
Fenêtre
?
Adobe PDF
Tapez une question
Arial
9
définition
Mode Formulaire
FILT
NUM
Démarrer
MSN H…
C:\Mé…
Docum…
Intégr…
carrier…
MédiT…
Reche…
Résult…
fiche-…
10:38

Acception
cutané 1d

Type Dérivé de nom

Domaine Médical

Base peau

Équivalent anglais

Paraphrase
NomT qui est effectué sur la peau

Équivalent coréen

NomT

- abrasion
- biopsie
- capitonnage
- examen
- incision
- moucheture
- prélèvement
- résection
- test

Enr : 1

Contexte(s)

En cas de forme cutanée, une biopsie cutanée peut être effectuée et analysée en immunohistochimie.

Source risquesbactério1.txt

L'intradermoréaction à la tuberculine (test de Mantoux) est un test cutané qui explore la réaction d'hypersensibilité retardée induite par les antigènes mycobactériens […].

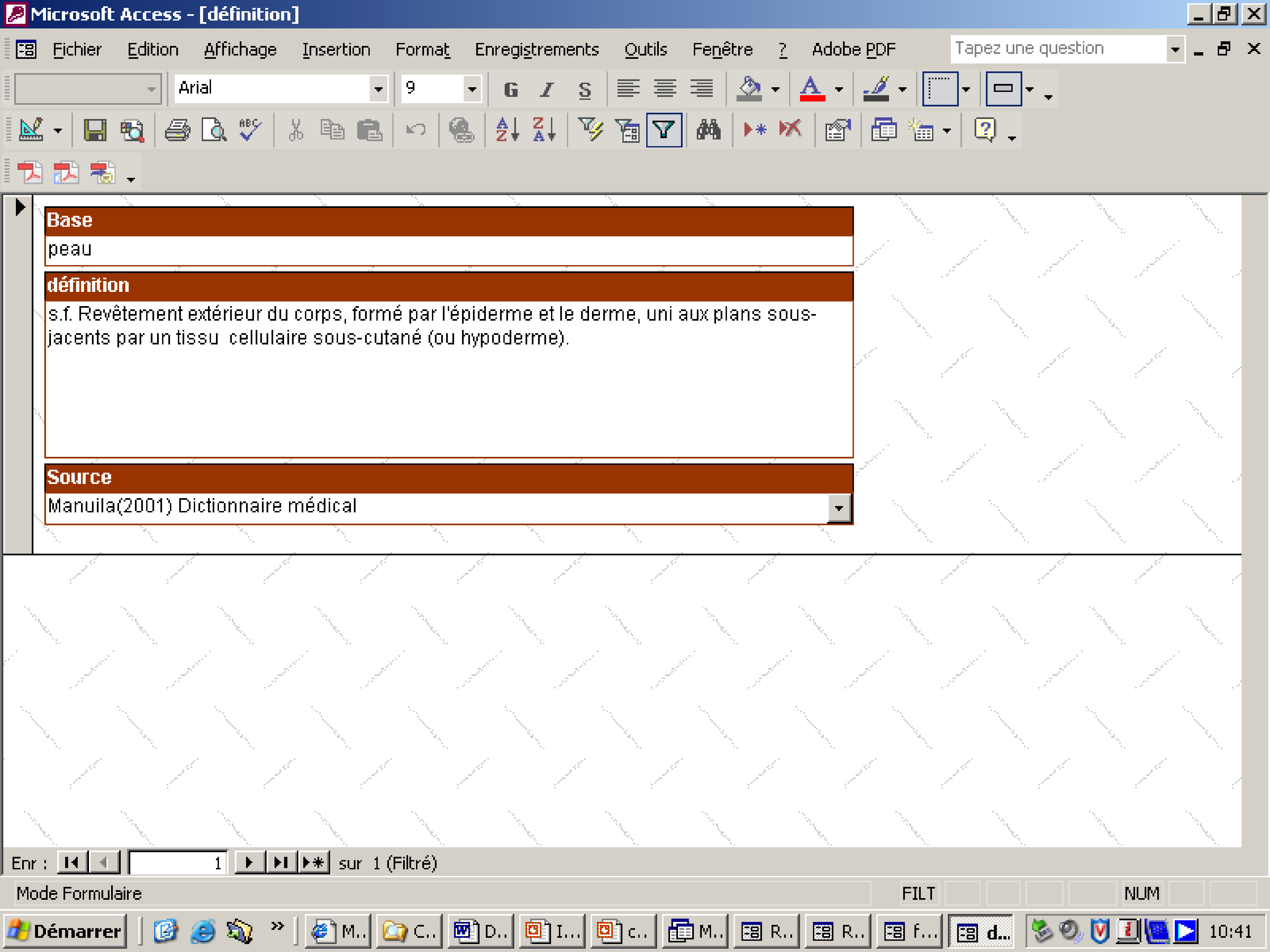
Source tubercul3.txt

En l'absence d'anomalies macroscopiques, quatre fragments par sein bénéficiaient systématiquement d'une fixation et d'une coloration selon les méthodes histologiques habituelles : il s' agissait d' un prélèvement cutané et de trois prélèvements glandulaires.

Enr : 1 sur 3

Enr : 1 sur 1 (Filtré)

Encoding of the data (7) BNoun definition


Microsoft Access - [définition]
Fichier
Edition
Affichage
Insertion
Format
Enregistrements
Outils
Fenêtre
?
Adobe PDF
Tapez une question
Arial
9
FILT
NUM
Démarrer
10:41

Base

peau

définition

s.f. Revêtement extérieur du corps, formé par l'épiderme et le derme, uni aux plans sous-jacents par un tissu cellulaire sous-cutané (ou hypoderme).

Source

Manuila(2001) Dictionnaire médical

Enr : 1 sur 1 (Filtré)

Mode Formulaire

Conclusion

The use of terminological tools presenting adjectives based on corpora studies allows the user to:

1. Aquire the semantic information carried by the adjective;
2. Better understand the sentence structure used in specialized languages;
3. Become familiar with lexical, syntactic and stylistic caracteristics, such as idioms and collocates.

Future research opportunities (1)

The development of a paraphrase system permits the observation of regularities among the semantic classes for a common relation.

Example

infection virale ‘viral infection’ =
infection qui est causée par un virus
‘infection which is caused by a virus’
inflammation bactérienne ‘bacterial inflammation’ =
inflammation qui est causée par une bactérie
‘inflammation which is caused by bacteria’

<état physiologique> qui est causé par <agent pathologique>
‘<physiological state> which is caused by a <pathological agent>’

Future research opportunities (2)

1- Build a model that works like the paraphrase system based on semantic classes.

2- Use the model to produce descriptions semi-automatically in French and in English (add paraphrases if necessary).

3- Develop the English version of DATUM using the model and find equivalents based on the paraphrases used to describe adjectives.