



TC/50/14  
 ORIGINAL : anglais  
 DATE : 13 mars 2014

**UNION INTERNATIONALE POUR LA PROTECTION DES OBTENTIONS VÉGÉTALES**  
 Genève

**COMITÉ TECHNIQUE**

**Cinquantième session**  
**Genève, 7-9 avril 2014**

DÉNOMINATIONS VARIÉTALES

*Document établi par le Bureau de l'Union*

*Avertissement : le présent document ne représente pas les principes ou les orientations de l'UPOV*

1. L'objet du présent document est de donner un aperçu général des faits nouveaux concernant la révision des "Notes explicatives concernant les dénominations variétales en vertu de la Convention UPOV" (document UPOV/INF/12/4), l'élaboration d'un moteur de recherche des similitudes pour l'UPOV à des fins de dénomination variétale et les domaines de coopération possibles entre la Commission internationale de nomenclature des plantes cultivées de l'Union internationale des sciences biologiques (Commission de l'UISB), la Commission de nomenclature et d'enregistrement des cultivars de la Société internationale de la science horticole (Commission de l'ISHS) et l'UPOV.

TABLE DES MATIÈRES

I.	RÉVISION ÉVENTUELLE DU DOCUMENT UPOV/INF/12 "NOTES EXPLICATIVES CONCERNANT LES DÉNOMINATIONS VARIÉTALES EN VERTU DE LA CONVENTION UPOV" .....	2
	<i>Travaux en cours du CAJ-AG sur les dénominations variétales .....</i>	<i>2</i>
	<i>Proposition sur les orientations concernant les dénominations variétales .....</i>	<i>2</i>
II.	POSSIBILITÉ D'ÉLABORATION D'UN MOTEUR DE RECHERCHE DES SIMILITUDES POUR L'UPOV AUX FINS DES DÉNOMINATIONS VARIÉTALES .....	3
	<i>Informations générales .....</i>	<i>3</i>
	<i>Création d'un groupe de travail .....</i>	<i>4</i>
III.	FAITS NOUVEAUX CONCERNANT LES DOMAINES DE COOPÉRATION POSSIBLES AVEC LA COMMISSION DE L'UISB ET LA COMMISSION DE L'ISHS.....	5
ANNEXE : PROCÉDURE DE RECHERCHE OCVV		

2. Les abréviations ci-après sont utilisées dans le présent document :

CAJ : Comité administratif et juridique

CAJ-AG : Groupe consultatif du Comité administratif et juridique

TC : Comité technique

Commission de l'UISB : Commission de l'Union internationale des sciences biologiques

Commission de l'ISHS : Commission de la Société internationale de la science horticole

I. RÉVISION ÉVENTUELLE DU DOCUMENT UPOV/INF/12 “NOTES EXPLICATIVES CONCERNANT LES DÉNOMINATIONS VARIÉTALES EN VERTU DE LA CONVENTION UPOV”

Travaux en cours du CAJ-AG sur les dénominations variétales

3. À sa huitième session tenue à Genève les 21 et 25 octobre 2013, le CAJ-AG a examiné le document CAJ-AG/13/8/6 “Matters concerning variety denominations”.

4. Le CAJ-AG a approuvé l'élaboration d'orientations concernant une demande d'obtenteur visant à changer une dénomination variétale enregistrée dans des cas autres que ceux dans lesquels la dénomination variétale a été radiée après l'octroi du droit d'obtenteur au motif que cette demande devrait être rejetée. Toutefois, le CAJ-AG a reconnu que des changements seraient justifiés dans les situations suivantes :

a) s'il était constaté l'existence d'un droit antérieur concernant la dénomination, qui aurait conduit au refus de cette dénomination (voir l'article 20.4) et 7) de l'Acte de 1991, l'article 13.4) et 7) de l'Acte de 1978 et la note 7 du document UPOV/INF/12/4);

b) si la dénomination ne convenait pas parce qu'elle était contraire aux dispositions de l'article 20.2) de l'Acte de 1991 et de l'article 13.2) de l'Acte de 1978; et

c) si la dénomination était par la suite refusée dans un autre membre de l'Union et si, à la demande de l'obtenteur, le service consentait à accepter la dénomination enregistrée dans cet autre membre de l'Union.

5. Le CAJ-AG est convenu que des orientations additionnelles devraient être envisagées dans le cadre d'une révision éventuelle des “Notes explicatives concernant les dénominations variétales en vertu de la Convention UPOV” (document UPOV/INF/12/4) (voir les paragraphes 69 à 71 du document CAJ-AG/13/8/10 “Compte rendu”).

6. Le rapport ci-dessus sur les travaux en cours du CAJ-AG consacrés aux dénominations variétales sera communiqué au CAJ à sa soixante-neuvième session qui se tiendra à Genève le 10 avril 2014 (voir les paragraphes 56 à 59 du document CAJ/69/2 “Élaboration de matériel d'information sur la Convention UPOV”).

Proposition sur les orientations concernant les dénominations variétales

7. Une version d'essai du cours d'enseignement à distance “Examen des demandes de droit d'obtenteur” (DL-305) en anglais a eu lieu du 11 novembre au 15 décembre 2013, avec la participation de neuf experts qui sont des formateurs pour le cours DL-205 agissant en qualité d'étudiants. L'observation suivante a été reçue d'un des étudiants à propos du paragraphe 2.3.3.a)i) du document UPOV/INF/12/4 (elle est reproduite ci-dessous par souci de commodité) :

“2.3.3 Identité de la variété

“a) De manière générale, une différence d'une seule lettre ou d'un seul chiffre peut être considérée comme susceptible d'induire en erreur ou de prêter à confusion quant à l'identité de la variété, sauf lorsque :

“i) la différence d'une lettre permet d'obtenir une différence visuelle ou phonétique nette, par exemple lorsqu'il s'agit d'une lettre au début d'un mot :

“Exemple 1 : en anglais, ‘Harry’ et ‘Larry’ ne prêtent pas à confusion; mais ‘Bough’ et ‘Bow’ peuvent prêter à confusion (phonétiquement);

“Exemple 2 : en japonais et en coréen, il n'y a pas de différence entre les sons ‘L’ et ‘R’, ce qui signifie que ‘Lion’ et ‘Raion’ se prononcent de la même façon alors qu'ils sont bien distincts pour les anglophones; [...]”

Observation :

“La sous-rubrique i), à savoir la différence d’une lettre permet d’obtenir une différence visuelle ou phonétique nette, par exemple lorsqu’il s’agit d’une lettre au début d’un mot, est suivie de l’exemple 1. Le cas de ‘Harry’ et ‘Larry’ illustre bien l’argument avancé. Ce qui me préoccupe, c’est celui de ‘Bough’ et ‘Bow’, deux mots qui pourraient être manifestement pris l’un pour l’autre pour des raisons phonétiques. Ils ne sont pas cependant un exemple de dénominations qui pourraient être prises l’une pour l’autre à cause de la différence d’une lettre – et dire qu’elles le sont pourrait être une source de distraction et de confusion pour les étudiants.

“Le cas de ‘Bough’ et ‘Bow’ pourrait-il être mieux couvert par une sous-rubrique traitant de la confusion phonétique? Je suppose que ma critique vise le document UPOV/INF/12 ainsi que le libellé du module.”

8. À sa soixante-neuvième session, qui se tiendra à Genève le 10 avril 2014, le CAJ sera invité à se demander s’il convient de modifier le document UPOV/INF/12, comme indiqué dans le paragraphe 7 du présent document.

9. Les observations du TC, à sa cinquantième session, seront communiquées au CAJ à sa soixante-neuvième session.

*10. Le TC est invité*

*a) à prendre note des travaux en cours du CAJ-AG sur l’élaboration d’orientations relatives aux dénominations variétales, telles qu’elles figurent dans les paragraphes 3 à 6 ci-dessus;*

*b) à prendre note que, à sa soixante-neuvième session, qui se tiendra à Genève le 10 avril 2014, le CAJ sera invité à se demander s’il convient de modifier le paragraphe 2.3.3.a)i) du document UPOV/INF/1, comme indiqué dans le paragraphe 7 du présent document; et*

*c) à formuler, le cas échéant, des observations sur la révision éventuelle du document UPOV/INF/12.*

## II. POSSIBILITÉ D’ÉLABORATION D’UN MOTEUR DE RECHERCHE DES SIMILITUDES POUR L’UPOV AUX FINS DES DÉNOMINATIONS VARIÉTALES

### Informations générales

11. À sa soixante-septième session tenue à Genève le 21 mars 2013, le CAJ a suivi un exposé de la délégation de l’Union européenne sur l’expérience de l’Office communautaire des variétés végétales (OCVV) concernant l’utilisation de son moteur de recherche des similitudes dans les dénominations lors de l’examen des dénominations proposées. Au cours de l’exposé, l’OCVV a proposé d’envisager la possibilité d’élaborer un moteur de recherche des similitudes pour l’UPOV aux fins des dénominations variétales, qui pourrait s’inspirer du moteur de recherche de l’OCVV<sup>1</sup>. Le CAJ s’est félicité de l’offre de l’OCVV et il est convenu d’inscrire un point à l’ordre du jour pour examiner cette proposition à sa soixante-huitième session, en octobre 2013 (voir les paragraphes 49 et 50 du document CAJ/67/14 “Compte rendu des conclusions”).

12. L’onglet Recherche de dénomination de la base de données sur les variétés végétales (base de données PLUTO) (<https://www3.wipo.int/pluto/user/fr/index.jsp>) propose à l’heure actuelle les modes de recherche ci-après, pour trouver des dénominations similaires :

<sup>1</sup> Le facteur de similarité a été élaboré par le *Groupe français d’étude et de contrôle des variétés et des semences* (GEVES) et l’Office communautaire des variétés végétales (OCVV) de l’Union européenne.

- “a) Facteur de similarité [moteur de recherche de l’OCVV] Ce mode permet d’analyser la dénomination que vous avez saisie en fonction d’une combinaison de facteurs tels que des lettres en commun, les longueurs relatives des mots et les positions des lettres communes. Il s’agit de la méthode de comparaison la plus complexe et la recherche peut prendre quelques secondes à effectuer. Le facteur de similarité a été élaboré par le GEVES (France) et l’Office communautaire des variétés végétales (OCVV). Veuillez toutefois noter que les résultats de la recherche à l’aide du facteur de similarité dans la base de données PLUTO nécessitent une interprétation et ne donnent pas de garantie quant à l’éligibilité des dénominations variétales, qui doit être décidée par l’administration à laquelle sont adressées les demandes de droits d’obtenteur.
- Une explication détaillée de l’analyse figure dans l’annexe du présent document.
- “b) Flou Ce mode permet de rechercher les dénominations contenant des mots qui s’écrivent comme les termes que vous avez saisis à un ou deux caractères près. Il est semblable à la méthode de mise en correspondance Flou de l’onglet Recherche de termes.
- “c) Phonétique Ce mode permet de rechercher les dénominations contenant des mots qui ressemblent phonétiquement aux termes que vous avez saisis. Il est semblable à la méthode de mise en correspondance Phonétique de l’onglet Recherche de termes.
- “d) Contient Ce mode permet de rechercher les dénominations contenant des mots qui comprennent la même série de lettres que les termes que vous avez saisis. Il est semblable à la méthode de mise en correspondance Contient de l’onglet Recherche de termes.
- “e) Commence Ce mode permet de rechercher les dénominations contenant des mots qui commencent par la même série de lettres que les termes que vous avez saisis. Il est semblable à la méthode de mise en correspondance Commence de l’onglet Recherche de termes.
- “f) Finit Ce mode permet de rechercher les dénominations contenant des mots qui finissent par la même série de lettres que les termes que vous avez saisis. Il est semblable à la méthode de mise en correspondance Finit de l’onglet Recherche de termes”.

13. Lors de discussions préliminaires avec le Bureau de l’Union sur la façon d’élaborer un moteur de recherche des similitudes pour l’UPOV aux fins des dénominations variétales, l’OCVV a précisé que toutes les options devraient être envisagées, et que, au vu des avancées dans le domaine des technologies de l’information, l’outil le plus efficace ne serait pas nécessairement basé sur le moteur de recherche de l’OCVV. La préoccupation première devrait être d’élaborer un outil que tous les membres de l’UPOV puissent utiliser, afin de limiter les divergences dans les décisions relatives à l’éligibilité.

#### Création d’un groupe de travail

14. À sa soixante-huitième session tenue à Genève le 21 octobre 2013, le CAJ a examiné le document CAJ/68/9 “Possibilité d’élaboration d’un outil de recherche de l’UPOV de similarité aux fins de la dénomination variétale” et approuvé la création d’un groupe de travail chargé d’élaborer des propositions relatives à un outil de recherche de l’UPOV de similarité aux fins de la dénomination variétale, comme proposé dans les paragraphes 4 à 7 du document CAJ/68/9, comme suit (voir le paragraphe 40 du document CAJ/68/10 “Compte rendu des conclusions”) :

“Ce groupe sera composé comme suit :

- “a) examinateurs de dénominations en provenance de membres de l’Union (trois à six experts);

- “b) Service des bases de données mondiales de l’OMPI (responsable de la base de données PLUTO);
- “c) Office communautaire des variétés végétales (OCVV) de l’Union européenne; et
- “d) Bureau de l’Union.

“Le programme de travail du groupe de travail sera établi par le groupe de travail lui-même; cependant, il est prévu que la première étape consiste à examiner les modes de recherche déjà disponibles dans l’onglet Recherche de dénomination de la base de données PLUTO, notamment le facteur de similarité (moteur de recherche de l’OCVV), ainsi qu’à examiner les modes de recherche utilisés dans d’autres situations (p. ex. en lien avec les marques) susceptibles de fournir une autre base à un outil de recherche de l’UPOV de similarité.

“L’examen de la pertinence des modes de recherche tiendra compte, en particulier, du document UPOV/INF/12 “Notes explicatives concernant les dénominations variétales en vertu de la Convention UPOV”. À cet égard, le groupe de travail devra se référer au CAJ pour obtenir des indications supplémentaires si ses travaux indiquent qu’un examen du document UPOV/INF/12 est nécessaire à l’élaboration d’un outil de recherche de l’UPOV de similarité efficace.

“Les réunions du groupe de travail seront accueillies et présidées par le Bureau de l’Union à Genève. Les réunions ne seront pas organisées de sorte à coïncider avec les sessions de l’UPOV et la participation électronique des examinateurs de dénominations et de l’OCVV sera prévue. Les propositions élaborées par le groupe de travail seront présentées au CAJ ainsi qu’au Comité technique (TC), et ces comités recevront un bref compte rendu des réunions du groupe de travail.”

15. À sa soixante-huitième session, le CAJ a pris note de la suggestion de la délégation de l’Union européenne tendant à inclure dans ce groupe de travail des examinateurs de dénominations en provenance de l’Espagne et des Pays-Bas, et indiqué qu’il importait de s’assurer que les experts couvrent suffisamment les aspects linguistiques des dénominations variétales (voir le paragraphe 41 du document CAJ/68/10 “Compte rendu des conclusions”).

16. À sa soixante-huitième session, le CAJ est convenu que les membres et observateurs devaient être encouragés à faire des suggestions sur les questions relatives aux tâches du groupe de travail (voir le paragraphe 42 du document CAJ/68/10 “Compte rendu des conclusions”). La première réunion du groupe de travail sera organisée en juin/juillet 2014.

17. Les observations du TC, à sa cinquantième session, seront communiquées au CAJ à sa soixante-neuvième session.

*18. Le TC est invité*

*a) à prendre note du rapport sur la possibilité d’élaboration d’un outil de recherche de l’UPOV de similarité aux fins de la dénomination variétale, qui figure dans la section II du présent document; et*

*b) à formuler, le cas échéant, des observations sur la possibilité d’élaboration d’un outil de recherche de l’UPOV de similarité aux fins de la dénomination variétale.*

III. FAITS NOUVEAUX CONCERNANT LES DOMAINES DE COOPÉRATION POSSIBLES AVEC LA COMMISSION DE L’UISB ET LA COMMISSION DE L’ISHS

19. Les informations générales relatives à cette question se trouvent dans les paragraphes 2 à 25 du document TC/49/8 “Dénominations variétales”.

20. À sa quarante-huitième session, tenue à Genève du 26 au 28 mars 2012, le TC a pris note du rapport de la délégation du Japon selon lequel la Commission de l’UISB était en train d’entreprendre la révision du Code international de nomenclature des plantes cultivées (CINPC) et qu’elle ferait des propositions à cette commission en 2013. Il est convenu que le Bureau de l’Union devrait contacter le CINPC pour expliquer les conseils donnés par l’UPOV dans le document UPOV/INF/12 “Notes explicatives concernant

les dénominations variétales en vertu de la Convention UPOV” (voir le paragraphe 90 du document TC/48/22 “Compte rendu des conclusions”).

21. À sa quarante-neuvième session, tenue à Genève du 18 au 20 mars 2013, le TC a examiné le document TC/49/8 “Dénominations variétales” et pris note des faits nouveaux concernant les domaines potentiels de coopération entre l’UPOV, la Commission de l’UISB et la Commission de l’ISHS), tels qu’ils figurent dans les paragraphes 24 et 25 du document TC/49/8 (voir le paragraphe 89 du document TC/49/41 “Compte rendu des conclusions”).

22. À sa soixante-huitième session, tenue à Genève le 21 octobre 2013, le CAJ a examiné le document CAJ/68/5 “Dénominations variétales”. Il a pris note que, le 19 juillet 2013, à Beijing (Chine), le Bureau de l’Union avait participé en qualité de conférencier au sixième Colloque international sur la taxonomie des plantes cultivées (ISTCP 2013) organisé par l’Université forestière de Beijing et le Jardin botanique de Beijing, sous les auspices de la Société internationale de la science horticole (ISHS). Le Bureau de l’Union avait donné des indications sur les conseils fournis par l’UPOV en ce qui concerne les dénominations variétales (voir le paragraphe 22 du document CAJ/68/10 “Compte rendu de conclusions”).

23. Le CAJ a pris note que, les 20 et 21 juillet, à Beijing également, le Bureau de l’Union avait participé en qualité d’observateur aux réunions de la Commission de l’UISB. Lors de ces réunions, la Commission de l’UISB avait examiné des propositions tendant à modifier la huitième édition du Code international de nomenclature des plantes cultivées (CINCP). Ces propositions ont été publiées en juillet 2013 dans le volume 7 du journal “Hanburyana” <http://www.rhs.org.uk/Plants/RHS-Publications/Journals/Hanburyana/Hanburyan-issues/Volume-7--June-2013>. Les propositions approuvées par la Commission de l’UISB figureront dans la neuvième édition du CINCP qui devrait être publiée en 2014.

24. À cette occasion, la Commission de l’UISB a proposé de constituer un groupe de travail, auquel participerait l’UPOV, dans les premières phases des travaux préparatoires de la dixième édition du CINCP. En marge des réunions de Beijing, le Bureau de l’Union s’était entretenu de manière informelle avec Mme Janet Cubey, présidente de la Commission de l’ISHS, afin d’examiner de quelle manière l’harmonisation des classes de dénomination pourrait être améliorée. Il a été suggéré d’étudier les domaines de coopération possibles en ce qui concerne les classes de dénomination au sein du groupe de travail proposé pour la dixième édition du CINCP.

25. Les travaux préparatoires concernant la dixième édition du CINCP ont eu lieu à une réunion des membres de la Commission de l’UISB le 4 mars 2014. Un rapport verbal sur les faits nouveaux pertinents sera remis au TC à sa cinquantième session.

*26. Le TC est invité à prendre note des faits nouveaux concernant les domaines de coopération possibles entre la Commission internationale de nomenclature des plantes cultivées de l’Union internationale des sciences biologiques (Commission de l’UISB), la Commission de nomenclature et d’enregistrement des cultivars de la Société internationale de la science horticole (Commission de l’ISHS) et l’UPOV, comme indiqué dans la section III du présent document.*

[L’annexe suit]



**OCVV**  
Office communautaire des variétés  
végétales

## PROCÉDURE DE RECHERCHE

### 1. Généralités

En conclusion de la phase d'étude du projet présenté au Conseil d'administration en novembre 2003, l'office a proposé, dans un premier temps, d'acquérir les logiciels nationaux, de les adapter à la base de données centralisée et de les exécuter pour chaque test. Les risques de rater une dénomination similaire seraient ainsi très limités.

À plus long terme, l'élaboration d'un logiciel de l'OCVV était envisagée, avec la possibilité de mettre au point des fonctionnalités linguistiques.

En pratique, les spécifications du logiciel français ont été choisies pour servir de base à la mise en œuvre de la procédure de recherche dans la base de données de l'OCVV.

### 2. Règles établissant une distinction suffisante entre deux dénominations variétales

L'une des dispositions du Règlement de base instituant un régime de protection communautaire des obtentions végétales prévoit qu'une dénomination variétale ne doit pas être identique à une dénomination variétale ou pouvoir être confondue avec une dénomination variétale sous laquelle une autre variété de la même espèce ou d'une espèce voisine est enregistrée.

Cette règle a été interprétée dans les principes directeurs du Conseil d'administration de l'OCVV sur les dénominations variétales :

- ♣ Une différence d'une seule lettre ou d'un seul chiffre, ou une différence d'accentuation, devrait généralement être considérée comme pouvant porter à confusion.
- ♣ Des différences de deux lettres ou plus ne devraient généralement pas être considérées comme pouvant porter à confusion, sauf lorsque les mêmes lettres sont simplement juxtaposées.
- ♣ En outre, une dénomination variétale ne devrait pas donner l'impression erronée que la variété est liée à une autre variété ou dérivée d'une autre variété.

Le programme de recherche visera à identifier des dénominations de la même classe dans la base de données qui pourraient être incompatibles avec une dénomination proposée.

### 3. Procédure de recherche

Les tests sont réalisés par un programme interne de la base de données ORACLE sur le serveur de l'OCVV (meilleurs résultats).

Pour réaliser un test, le programme d'interface (site Web, etc.) exécute une procédure appelée :

## TESTDENOMINATION

La dénomination visée par le test et le code de l'espèce à laquelle appartient la variété sont entrés comme paramètres par le programme d'interface.

La procédure renvoie l'identifiant du test réalisé vers l'interface (colonne Testid). Cet identifiant permet de visualiser les deux tableaux qui constituent les résultats du test : TESTS et TESTRESULTS.

Le tableau TESTS contient les données générales relatives au test : date du test, identifiant de la personne qui a demandé le test, dénomination, code de l'espèce, code de la classe ou du genre, temps d'exécution de l'ordinateur, mots exclus, message d'erreur, etc.

Le tableau TESTRESULTS contient les listes des dénominations qui ont été considérées comme similaires par la procédure TESTDENOMINATION. Chaque dénomination similaire est associée à un indice de similarité.

### 4. Description de la procédure TESTDENOMINATION

#### Contrôle d'entrée

Le code de l'espèce doit figurer dans la base de données des dénominations.

La dénomination ne doit pas compter plus de 100 caractères (espaces inclus).

Caractères exclus : signes de ponctuation et /-\_' . Les caractères accentués ne sont pas autorisés.

Les dénominations composées de plusieurs mots sont autorisées, dans la limite de quatre mots. Les mots doivent être séparés par un espace.

#### a) Première opération : division de la dénomination de l'entrée en mots distincts

La dénomination visée par le test est divisée en mots élémentaires. Les espaces, considérés comme des séparateurs, sont supprimés.

Les caractères standards non latins (caractères accentués, etc.) sont remplacés par des caractères standards latins. Toutes les lettres sont transformées en lettres capitales. Voir le tableau de conversion.

*Exemple : Déjà devient DEJA*

Les lettres doubles deviennent des lettres simples.

*Exemple : HELLO devient HELO.*

Chaque mot élémentaire est confronté à la liste de mots exclus du test (p. ex. color YELLOW, RED, PURPLE, etc.)

Les mots élémentaires figurant dans cette liste sont exclus du test de similarité.

Dans la description ci-après, les mots élémentaires obtenus sont appelés "Mots testés" (*Words Tested – WT*)

Exemple : dénomination "Tānau TARI YELLOW"

Le mot *yellow* est exclu des tests. Les mots pris en compte (WT) sont les suivants :

- TANAU,
- TARI,
- TANAUTARI,
- TARIYELLOW,
- TANAUTARIYELLOW.



## b) Seconde opération : établissement d'une liste de mots de référence dans la classe

Le programme recherche la classe ou le genre associé à l'espèce pour déterminer la portée de la recherche de similitude.

Pour les dénominations enregistrées en tant que codes, une chaîne unique de caractères est prise en compte, sans espace.

Le logiciel établit une liste de tous les mots élémentaires appartenant aux dénominations des variétés de la classe.

Tous les mots comptant plus de 3 caractères que le WT le plus long et moins de 3 caractères que le WT le plus court ne sont pas pris en compte.

Tous les mots de référence appartenant à la liste susmentionnée de mots exclus des tests (= color YELLOW, RED, PURPLE, etc.) sont exclus du test de similitude.

Comme précédemment, les lettres doubles des mots de référence sont transformées en lettres simples.

Dans la description ci-après, les mots élémentaires inclus dans cette liste sont appelés "Mots de référence" (*Words of reference* – WR).

## c) Les principes du test de similitude

Pour chaque WT, la procédure SIMILARITYTEST calcule un indice de similitude par rapport à chaque WR inclus dans la liste de dénominations précédemment établie.

Les WR sont classés en fonction de la valeur de l'indice de similitude. Tous les WR dont l'indice de similitude est supérieur à un seuil prédéfini sont exclus des résultats.

Exemple :

Dénomination testée "Tānau TARI YELLOW"

La liste de mots de référence contient 10 000 mots élémentaires.

TANAU, TARI, TANAUTARI, TARIYELLOW et TANAUTARIYELLOW sont testés par rapport à ces 10 000 mots de référence. Cinquante mille tests sont réalisés (5 mots x 10 000 WR).

## 5. Description détaillée du test de similitude

Les étapes ci-après sont mises en œuvre pour chaque paire (WT, WR).

Filtre préliminaire : dans la liste de WR, tous les mots comptant plus de 3 caractères que le WT et moins de 3 caractères que le WR ne sont pas pris en compte pour les calculs suivants.

### Première étape : calcul de $Ki2$

Formule :  $Ki2 = \text{somme}(di)^2 / (\text{longueur}(WT)-1)(\text{longueur}(WR)-1)$

Où  $di$  = différence de nombre de lettres entre le mot testé et le mot de référence. Toutes les lettres des deux mots sont prises en compte.

Exemple :

WT : **ALADIN** – 6 caractères.

Si l'on compare ce WT avec la chaîne de caractères existante : **DYLAN** (5 caractères)

	A	L	D	I	N	Y
<b>ALADIN</b>	2	1	1	1	1	0
<b>DYLAN</b>	1	1	1	0	1	1

$$\text{Chi2} = ((2-1)^2 + (1-1)^2 + (1-1)^2 + (1-0)^2 + (1-1)^2 + (0-1)^2) / (6-1)(5-1)$$

$$\text{Chi2} = (1 + 0 + 0 + 1 + 0 + 1)(5*4)$$

$$\text{Chi2} = 3/20$$

$$\text{Chi2} = 0,15$$

On conserve donc pour le calcul suivant tous les mots pour lesquels :

- $\text{Ki2} \leq 0,3$  et longueur du WT  $\geq 5$  caractères
- $\text{Ki2} \leq 0,4$  et longueur du WT = 4 caractères
- $\text{Ki2} \leq 0,5$  et longueur du WT < 4 caractères

## Deuxième étape : quatre calculs basés sur la sélection de la première étape

Au cours de cette deuxième étape, quatre calculs sont réalisés :

- Calcul du pourcentage de lettres communes
- Calcul du pourcentage de lettres NON communes
- Calcul du pourcentage de différence de longueur
- Calcul du coefficient de corrélation des rangs de Kendall

### Calcul du pourcentage de lettres communes (*common letter* = CL)

$$\text{CL} = 1 - (\text{nombre de lettres communes}) / (\text{longueur (WT)})$$

Exemple : ALADIN et DYLAN

$$\text{CL} = 1 - 4/6 = 0,33$$

CL est égal à 0 lorsque toutes les lettres figurent dans le WR.

Second exemple : BANANAS et BANS

Toutes les lettres de BANANAS figurent dans le mot BANS.

$$\text{CL} = 1 - 7/7 \Rightarrow \text{CL} = 0$$

### Calcul du pourcentage de lettres NON communes (*non-common letter* = NCL)

$$\text{NCL} = (\text{nombre de lettres du WR absentes dans le WT}) / (\text{longueur (WT)})$$

Exemple : ALADIN et DYLAN

$$\text{NCL} = 1/6 = 0,16$$

NCL est égal à 0 lorsque toutes les lettres du WR figurent dans le WT.

Second exemple : BANANAS et BANS

Toutes les lettres de BANS figurent dans le mot BANANAS.

$$\text{NCL} = 0/7 \Rightarrow \text{NCL} = 0$$

### Calcul du pourcentage de différence de longueur (*difference of length = DL*)

$DL = (\text{différence de longueur entre les 2 chaînes}) / (\text{longueur du WT})$

Exemple : BANANAS et ANANAS

$DL = 1/7$

DL est égal à 0 lorsque les deux mots sont de la même longueur.

### Calcul du coefficient de corrélation des rangs de Kendall (*Kendall correlation = KC*)

Formule :  $KC = 6 * \text{somme}(Di)^2 / N * (N^2 - 1)$

Où :

- Di est égal à la différence de position des lettres communes (Li) moins la différence de position des lettres précédentes. Si le WT contient plusieurs occurrences de la même lettre, les premières lettres sont utilisées comme référence pour la position. Si le WR contient plusieurs occurrences de la même lettre, on utilise les lettres les plus proches de la même lettre dans le WT.
- N est égal au nombre de lettres communes entre le WT et le WR

Exemple :

WT : **ALADIN** et **DYLAN**

	<b>A</b>	<b>L</b>	<b>D</b>	<b>N</b>
<b>ALADIN</b>	1	2	4	6
<b>DYLAN</b>	4	3	1	5
<b>Différence</b>	1-4 = -3	2-3 = -1	4-1 = 3	6-5 = 1
<b>Di</b>	-3-0 = -3	1-(-3) = 4	3-(-1) = 4	1-3 = -2

4 lettres en commun A, L, D et N

$$KC = 6 * ((-3)^2 + (4)^2 + (4)^2 + (-2)^2) / (4 * (4^2 - 1))$$

$$KC = 6 * (9 + 16 + 16 + 4) / 4 * 15$$

$$KC = 6 * 45 / 60$$

$$KC = 4,5$$

WT : **ALADIN** et **BALADIN**

	<b>A</b>	<b>L</b>	<b>D</b>	<b>I</b>	<b>N</b>
<b>ALADIN</b>	1	2	4	5	6
<b>BALADIN</b>	2	3	5	6	7
<b>Différence</b>	1-2 = -1	2-3 = -1	4-5 = -1	5-6 = -1	6-7 = -1
<b>Li</b>	-1-0 = -1	-1-(-1) = 0	-1-(-1) = 0	-1-(-1) = 0	-1-(-1) = 0

5 lettres en commun A, L, D, I et N

$$KC = 6 * ((-1)^2 + (0)^2 + (0)^2 + (0)^2 + (0)^2) / (5 * (5^2 - 1))$$

$$KC = 6 * (1) / 5 * 24$$

$$KC = 1 / 20$$

$$KC = 0,05$$

Dans cet exemple, on peut voir que si la même séquence de lettres figure dans les deux mots (ici LADIN), l'écart entre les deux séquences n'est pris en compte qu'une seule fois.

Si les deux mots WT et WR sont identiques, le coefficient de Kendall est égal à 0.

## Troisième étape : calcul de l'indice de similitude

Indice de similitude = KC + CL + NCL + DL

Un mot de référence est **sélectionné** si les scores sont inférieurs ou égaux à :

	<b>KC</b>	<b>CL</b>	<b>DL</b>	<b>Indice de similitude</b>
Longueur du mot recherché > 4 caractères	<= 1,5	<= 0,22	<= 1,5	< 1,2
Longueur du mot recherché = 4 caractères	<= 1,5	<= 0,25	<= 1,26	< 1,2
Longueur du mot recherché < 4 caractères	<= 1	<= 0,34	<= 1,0	< 1,2

Les dénominations de référence sont triées en fonction de l'indice de similitude, puis dans l'ordre alphabétique.

[Fin de l'annexe et du document]