

Aspects du traitement automatique du gallois

Johannes Heinecke

France Télécom, Division Recherche & Développement

2 avenue Pierre Marzin, F-22307 Lannion Cedex

johannes.heinecke@francetelecom.com

Mots-clefs : typologie, lexique, morphologie, syntaxe, ressources, corpora, TALN, gallois

Keywords: typology, lexicon, morphology, syntax, resources, corpora, computational linguistics, Welsh

Résumé Cet article décrit les ressources linguistiques (informatisées) du gallois, une langue celtique parlée par plus que 500.000 personnes au Pays de Galles (Royaume Uni). Après une introduction brève sur la situation actuelle du gallois, ses spécificités typologiques et des difficultés potentielles pour un traitement automatique sont présentées. Ensuite les ressources linguistiques électroniques disponibles et/ou envisageables sont discutées. Pour finir nous présentons quelques travaux dans le domaine du traitement automatique du gallois.

Abstract This article describes the (electronic) linguistic resources for Welsh, a Celtic Language spoken by more than half a million persons in Wales (UK). After a short introduction to the current situation of the Welsh language, its main typological features are presented, including potential difficulties for all areas of natural language processing. In a following section the electronically available linguistic resources are listed and discussed. Finally an eye is cast onto some work on Welsh within computational linguistics.

1 Introduction

Le gallois (Cymraeg) est, à côté de l'anglais, une des langues du Pays de Galles au Royaume Uni. C'est une langue celtique, plus précisément, une langue P-celtique insulaire.¹ Elle est la sœur du Breton (parlé en Bretagne en France) et du Cornouaillais, une langue disparue depuis la fin du XVIII^e siècle, qui était parlée au Cornwall (sud-ouest de la Grande Bretagne). Les autres langues celtiques (Q-celtique) sont l'irlandais, le gaélique de l'Ecosse et le manx (disparu récemment, mais toujours parlé par quelques personnes qui l'ont appris comme deuxième langue).

Dans cet article nous donnons un aperçu général sur le gallois et dans un cadre plutôt linguistique, ses caractéristiques typologiques. Dans une autre section nous présentons les ressources linguistiques électroniques existantes et leur utilité pour un traitement automatique. Nous finissons cet article avec une brève introduction aux travaux dans le domaine du TALN pour le gallois. Nous nous restreignons ici à un panorama général du gallois et son traitement automatique. Pour plus de détails voir les références bibliographiques.

2 La langue galloise

2.1 Situation actuelle

Depuis 1998² le gallois est reconnu comme langue officielle au Pays de Galles à côté de l'anglais. Elle est utilisée dans les administrations et enseignée ou la langue d'enseignement dans beaucoup d'écoles et dans l'université du Pays de Galles. Un l'organisme du gouvernement, le *Welsh Language Board*³ assure que l'utilisation de la langue galloise est possible dans tous les endroits de la société et fait la promotion de la langue. Selon le dernier recensement 2001, presque 21% de la population galloise (i.e. 582.400 personnes) parle le gallois.⁴ Par rapport au recensement antérieur (1991), cela signifie une augmentation d'environ 80.000 personnes (1991 : 18,7%). Grâce à l'enseignement obligatoire, même dans les régions plutôt anglophones (sud-est), le pourcentage des jeunes qui parle le gallois est en forte croissance (cf. tableau 1).

Actuellement le gallois n'est pas immédiatement menacé de disparaître, ceci grâce aux nouvelles structures mises en place pendant les dernières décennies afin de protéger et promouvoir l'utilisation du gallois. Par contre, le fait que quasiment tous les locuteurs du gallois sont bilingues (anglais) a des effets sur le vocabulaire et la syntaxe du gallois (parlé). On y trouve nombreux anglicismes et de constructions syntaxiques issues de l'anglais, comme la position finale de la proposition *o* « de » au lieu d'être avant les pronoms interrogatifs :

- (1) *lle wyt ti 'n dod o?* syntaxe correcte : *o le wyt ti 'n dod?*
 où es tu IMPF venir de de+SM SM-où es tu IMPF venir
 « d'où est-ce que tu viens ? » (cf. « where do you come **from** ? »)

¹Pour la distinction entre P-celtique et Q-celtique voir (Ball & Fife, 2002).

²Government of Wales Act 1998

³<http://www.bwrdd-yr-iaith.org.uk/>

⁴Voir l'information sur ce dernier recensement sur le site du *Welsh Language Board*.

Age	personnes	pourcentage
3-4	13.239	18,8 %
5-15	171.168	40,8 %
16-19	40.548	27,6 %
20-44	146.227	15,5 %
45-64	112.742	15,6 %
65-74	47.692	18,1 %
75+	50.752	21,1 %

TAB. 1 – résultats du recensement 2001 : personnes capables de parler le gallois

À l'exception des journaux quotidiens, il existe depuis longtemps une grande diversité d'hebdomadaires et de magazines spécialisées. La vie littéraire est en bonne santé avec plus de 500 nouvelles publications par an.⁵

2.2 Caractéristiques typologiques

Malgré le fait que les langues celtiques, et donc le gallois, sont des langues indo-européennes, elles montrent quelques traits typologiques qu'on ne trouve pas parmi les autres langues de la même famille de langues. La plupart de ces différences concerne toutes les langues celtiques insulaires. La description suivante est en partie spécifique pour le gallois. La plupart de ces traits typologiques ne pose aucun problème à un traitement automatique du langage, par contre les approches de base, développées souvent sur la typologie de l'anglais, allemand, français etc. peuvent être trop spécifiques pour traiter le gallois. Par exemple l'ordre des mots de base du gallois est VSO, i.e. le verbe prend toujours la première place dans une phrase.

Un phénomène très connu dans les langues celtiques sont les mutations des consonnes initiales des mots en fonction de leur fonction syntaxique. Le gallois connaît trois mutations, la lénition (*soft mutation*, SM), la nasalisation (*nasal mutation*, NM) et l'aspiration (*aspirate mutation*, AM).⁶ Par exemple, le complément d'objet direct est muté : *Gwelodd Ioan gi* « Ioan a vu un chien ». La forme dite « radicale » du mot pour chien est *ci*, qui apparaît ici avec lénition. Certaines consonnes ne mutent pas, par contre l'aspiration des voyelles implique un *h* préfigé : *afal* « une pomme » mais *ei hafal* « sa pomme (à elle) ». Il y a des cas où seulement la mutation peut désambiguïser le syntagme : lénition *ei dŷ* « sa maison (à lui) » vs. aspiration : *ei thŷ* « sa maison (à elle) » ou encore *daeth o* « il est venu » vs (*a*)⁷ *ddaeth o ?* « est-ce qu'il est venu ? ».

Toutes les langues celtiques connaissent des prépositions fléchies (nombre, personne et genre) devant des pronoms : *gennyf i* « avec moi », *ganddo fo* « avec lui », mais *gan y dyn* « avec l'homme ». Ces prépositions fléchies, l'ordre basique des mots, le comportement de l'article

⁵Welsh Books Council, <http://www.wbc.org.uk/>

⁶Voir Ball & Müller, 1992; Borsley, 1999.

⁷La particule interrogative *a* n'est pas obligatoire et disparaît normalement dans la langue parlée et souvent dans le gallois écrit moins formel. Ce ne laisse que la mutation comme le seul trait qui indique l'interrogation.

défini qui rend défini une construction génitive ont provoqué depuis le début du XIX^e siècle, les hypothèses que les langues celtiques insulaires ont été en contact avec des locuteurs des langues afro-asiatiques, notamment les langues sémitiques, qui sont les seules langues qui partagent ces phénomènes assez étonnants (Pokorny, 1927-1930; Wagner, 1959; Gensler, 1993; Vennemann, 2002).

Le gallois, notamment sa version parlée, possède un système verbal analytique complexe afin d'exprimer le temps et l'aspect (voir Heinecke, 1999). Contrairement aux autres langues indo-européennes (mais comme l'irlandais par exemple), le gallois utilise des particules (historiquement des prépositions) afin d'exprimer le temps sémantique (exemple 2) et/ou l'aspect (exemple 3) :

- (2) *Mae 'm brawd wedi cyrraedd*
 est mon frère nouveau arriver
 « mon frère vient d'arriver » littéralement « mon frère est après d'arriver »
- (3) *Roeddwn ni 'n canu trwy 'r nos*
 AFF-étions nous IMPF chanter pendant la nuit
 « nous chantions pendant toute la nuit »

Mis à part sa structure temporelle, l'accord entre le verbe et le sujet n'est complet que si le sujet est un pronom. Pour un sujet nominal, le verbe est toujours à la 3^e personne du singulier.

Au niveau syntaxique, il est à observer que le gallois n'a pas de diathèse (passif) mais il y a des formes impersonnelles (une septième forme en plus des 3 personnes singulier et pluriel) qui reprennent une partie du passif : Une expression comme *argraffwyd yng Nghymru* « imprimé au Pays de Galles » littéralement « on (l') a imprimé au Pays de Galles » est utilisée, quand on ne connaît pas l'agent (où le premier actant). Syntactiquement il ne s'agit pas d'un passif, car cette construction est similaire aux autres formes actives.⁸ Il correspond plutôt au *on* français ou *man* allemand. Dans la langue parlée les constructions avec l'auxiliaire *cael* « avoir, recevoir » font émerger quelque chose qui est plus comparable avec le passif français etc. :

- (4) *Mi gês i fy ngheni ym Mis Chwefror*
 AFF+SM SM-avoir-PRÉT-1SG je+NM NM-mon NM-naître en mois février
 « je suis né en février », littéralement « j'ai reçu mon naître en février »

Un autre phénomène syntaxique intéressant est le « fonction spreading ». Il s'agit d'une phrase coordonnée dont seulement le premier verbe est fléchi, les autres verbes étant des infinitifs (noms verbaux). La personne, le nombre, le temps et l'aspect du verbe initial sont implicitement sous-entendus pour les autres verbes⁹ :

- (5) *cododd Gwilym ei docyn a rhedeg am y trêen*
 acheter-3SG-PRÉT Gwilym son+SM SM-billet et courir-VN pour le train
 « Gwilym acheta son billet et courut pour le train » (Thomas, 1996)

⁸Sauf le fait que le COD indéfini n'est pas lénifié. Voir aussi Awbery, 1976.

⁹Voir plus dans le cadre de la LFG dans Sadler, 2003.

Enfin, syntaxiquement la possession et le complément d'objet direct ne sont pas différenciés : En fonction de la catégorie de sa tête, le pronom possessif¹⁰ exprime soit le possesseur soit l'objet :

(6) *mae o 'n fy ngweld (i)*
 est il IMPF mon+NM NM-voir je
 « il est en train de me voir »

mae hwn yn fy nhŷ (i)
 est ce PRED ma+NM maison (je)
 « c'est ma maison »

Une difficulté à résoudre est l'existence de deux standards *de facto*. Il existe une norme plus conservatrice du gallois qui est surtout utilisée dans les textes écrits (par exemple dans les journaux ou dans la littérature). Cette version du gallois utilise plus souvent des formes verbales synthétiques au lieu des formes analytiques : Par exemple dans le gallois plutôt formel le passé antérieur est exprimé par une temps grammatical nommé « Irrealis Tense »,¹¹ mais de préférence par une périphrase analytique dans les occasions moins formelles : *canasem* « nous aurions eu chanté » vs *mi fasen ni wedi canu*. De plus, dans le gallois plus formel, les particules verbales ne sont pas disparues ou confondues avec les racines verbales, mais restent telles quelles :

(7) *yr ydwyf i [wedi darllen y llyfr]*
 AFF être-1SG-PRÉS je après lire le livre
 « j'ai lu le livre »

vs

(8) *(ry)dwi [wedi darllen y llyfr]*
 être-1SG-PRÉS après lire le livre

À la lecture des grammaires descriptives de Williams, 1980 et celles de King, 1993 et Thorne, 1993 on pourrait avoir l'impression qu'il s'agit de deux langues proches mais différentes. Pour un traitement automatique il faut pourtant prévoir des grammaires (c'est à dire des ensembles de règles pour l'analyse syntaxique) qui permettent d'analyser tous ces phénomènes à la fois, car, notamment dans les textes moins formels, les deux types de syntagme peuvent être trouvés.¹²

Le gallois est une langue très bien documentée depuis son début. Il existe des nombreuses grammaires descriptives, qui couvrent, d'un côté l'aspect formel du gallois, de l'autre la langue plus spontanée.¹³ De plus, le gallois est l'objet d'étude pour la validation de plusieurs théories syntaxique, comme la Generative Transformational Grammar (Awbery, 1976), la théorie X-Bar

¹⁰C'est pour cette raison que dans les grammaires descriptives, ces pronoms s'appellent « pronoms dépendants » au lieu de « pronom possessif ». Les pronoms personnels en fonction de sujet sont nommés « pronoms indépendants ».

¹¹Heinecke, 1999.

¹²Ball *et al.*, 1988 et Jones, 1988.

¹³Par exemple : Morris-Jones, 1913; Morris-Jones, 1921; Williams, 1980; Humphreys, 1980; King, 1993; Thorne, 1993; Thomas, 1996.

(Rouveret, 1990, Roberts, 2004), la Lexical Functional Grammar (LFG, par exemple Sadler, 1998 et Sadler, 1999) ou encore la Head-Driven Phrase Structure Grammar (HPSG, Borsley, 1990). Le gallois est le sujet d'innombrables études typologiques (voir Heinecke, 1999). La recherche en linguistique comparative et des études sur les langues indo-européennes ont bien étudié l'histoire et le développement du gallois.

La description phonologique est aussi complète comme la présentation d'un standard parlé (Thomas, 1987) et l'atlas linguistique du gallois (Thomas, 1973) le montrent. Le gallois a deux groupes de dialectes principaux, qui se différencient au niveau phonologique (Thomas & Thomas, 1989) mais aussi morphologiquement et lexicalement. Pour des applications de reconnaissance vocale, deux modèles de langue sont exploités (voir ci-dessous).

3 Technologie de l'information pour le gallois

3.1 Utilisation dans l'informatique

Depuis la définition de l'encodage ISO-8859-14,¹⁴ les caractères de l'orthographe galloise sont tous accessibles. En plus de l'alphabet latin de base, le gallois connaît des accents circonflexes sur toutes les voyelles comme marqueur de longueur; le fait que *y* et *w* sont des graphèmes qui représentent des phonèmes vocaliques, nécessite des accents sur eux aussi : *Ŷ/ŷ* (xDE/xFE), *Ŵ/ŵ* (xD0/xF0). Un autre point à ne pas négliger sont les bigraphes *ch*, *dd*, *ff*, *ll*, *ng*, *ph*, *rh* et *th*, qui sont considérés comme étant des caractères simples. En plus, *ng* suit *g* et ne pas *n* dans l'alphabet. Donc pour un tri alphabétique il faut assurer que *didwyll* précède *didda*, *engyl* précède *ehangdeb* etc.

Depuis plusieurs années, les travaux de localisation des logiciels en gallois ont été menés. Récemment une version galloise du système d'exploitation Windows XP a été publiée. En ce qui concerne Linux, des logiciels bureautiques tels que OpenOffice, Mozilla etc. ont désormais une interface galloise.¹⁵ Le Canolfan Bedwyr¹⁶ de l'Université du Pays de Galles à Bangor est chargé de promouvoir et mener des projets afin d'obtenir des logiciels en gallois et pour le gallois (TALN), depuis 1999 le correcteur d'orthographe *CySill* est disponible.¹⁷ *CySill* était une nouveauté parce que c'était le premier correcteur d'orthographe qui maîtrisait les mutations.

3.2 Ressources électroniques

Les ressources linguistiques autres que celles du Canolfan Bedwyr sont rares. Le Dictionnaire de l'Université du Pays de Galles (*Geiriadur Prifysgol Cymru*, GPC), le dictionnaire gallois

¹⁴L'unicode, évidemment, couvre aussi la totalité des caractères qui sont nécessaire pour utiliser le gallois dans l'informatique.

¹⁵<http://www.meddal.org/>

¹⁶<http://www.bangor.ac.uk/ar/cb/>

¹⁷*CySill* fait partie du package logiciels *Cysgliad* qui contient en plus des dictionnaires terminologiques *Cysgeir*. *Cysgliad* est disponible au Canolfan Bedwyr.

de référence¹⁸. Un CD-ROM du GPC est envisagé prochainement. Plusieurs personnes, dont l'auteur, ont créé des dictionnaires accessibles sur l'internet,¹⁹ mais ils ne suffisent pas pour un traitement automatique du gallois, car il s'agit des listes de lemmes, sans modèles de flexions etc.

Malgré le fait que la tradition des corpus a commencé au début du xx^e siècle [Fynes-Clinton, 1913](#), il n'existe que très peu de corpus électroniques. Kevin Scannell de l'Université de St. Louis (Missouri, États-Unis) a implémenté des logiciels qui permettent d'établir des corpus pour les langues peu dotées à partir des documents trouvés sur le Web. Le corpus du gallois comprend 69.502 fichiers, avec environ 95.800.000 mots (septembre 2004).²⁰ Un autre corpus important est le *Cronfa Electronig o Gymru* (« Archive électronique du Pays de Galles ») qui a été compilé en 1994 à partir de textes littéraires, de journaux, et de documents scientifiques et commerciaux. Ce corpus est d'environ un million de mots ; il est étiqueté et contient une statistique des mots, leurs mutations et leurs catégories grammaticales ([Ellis et al., 2001](#)).

Un troisième corpus qui est en train de se constituer est le *Historical Corpus of the Welsh Language*.²¹ Dans ce projet, mené à l'Université de Cambridge, on veut créer un corpus des textes de la période 1500 à 1850. Son intention est plutôt de rendre disponible une base de données pour les études de linguistique historique comme par exemple les changements syntaxiques et le développement du vocabulaire. Ce corpus sera encodé en XML en utilisant le codage développé par le *Text Encoding Initiative*²² (TEI). Ce projet offre aussi une concordance dans le but des études philologiques. Pour l'instant (décembre 2004) ce corpus contient 30 textes (420.000 mot).

3.3 Traitement automatique du langage

À ce jour le traitement automatique du langage du gallois autre que la correction d'orthographe est quasiment inexistant. Aux Canolfan Bedwyr en collaboration avec le Irish Speech Group du Trinity College Dublin, un système de synthèse, développé à l'origine à l'Université d'Edimbourg ([Williams, 1999](#)), est réutilisé dans le projet *Welsh and Irish Speech Processing Resources* (WISPR).²³ Un corpus du gallois oral a été créé dans le cadre du projet *SpeechDat Cymru* à l'Université du Pays de Galles à Swansea ([Jones et al., 1998](#)). Ce corpus contient des phrases lues au téléphone par 2000 locuteurs de toutes les régions galloises. Chaque enregistrement est ensuite annoté manuellement.

Dans l'équipe Langues Naturelles de la Division Recherche de France Télécom nous avons

¹⁸[Thomas & Bevan, 1950-2002](#) : la première édition a été terminée en 2002 après autour de 80 ans de travail (qui avait commencé en 1921, bien avant la publication du premier fascicule) et contient environ 84.400 d'entrées (lemmes). Depuis, l'équipe du GPC travaille sur une réédition totale. Cf. aussi <http://www.aber.ac.uk/~gpcwww/>. Le dictionnaire inverse (anglais-gallois, [Griffiths & Jones, 1995](#)) est aussi basé sur une base de données dont la lettre *a* était accessible en ligne à http://www.swan.ac.uk/uwp/wa_index.htm

¹⁹Gallois et anglais : <http://www.cs.cf.ac.uk/fun/welsh/LexiconForms.html> ; gallois, anglais et catalan http://www.estelnet.com/catalunyacymru/catala/gbs_mynegai_1818.htm ; gallois et allemand (par l'auteur) : <http://perso.wanadoo.fr/heinecke/dict/cymraeg/>.

²⁰<http://borel.slu.edu/crubadan/>

²¹<http://www.mml.cam.ac.uk/ling/research/welshcorpuseng.html>

²²<http://www.tei-c.org/>

²³<http://www.bangor.ac.uk/ar/cb/wispr.php>

récemment adapté notre boîte à outils pour le TALN au gallois (écrit). Un petit lexique comportant les modèles flexionnels nécessaires et un ensemble de règles grammaticales (dans le cadre d’une syntaxe de dépendance) nous permet d’analyser le gallois et générer une représentation sémantique (phrase à analyser *mi fwytodd y llygoden gaws* « la souris a mangé le fromage », figure 1).

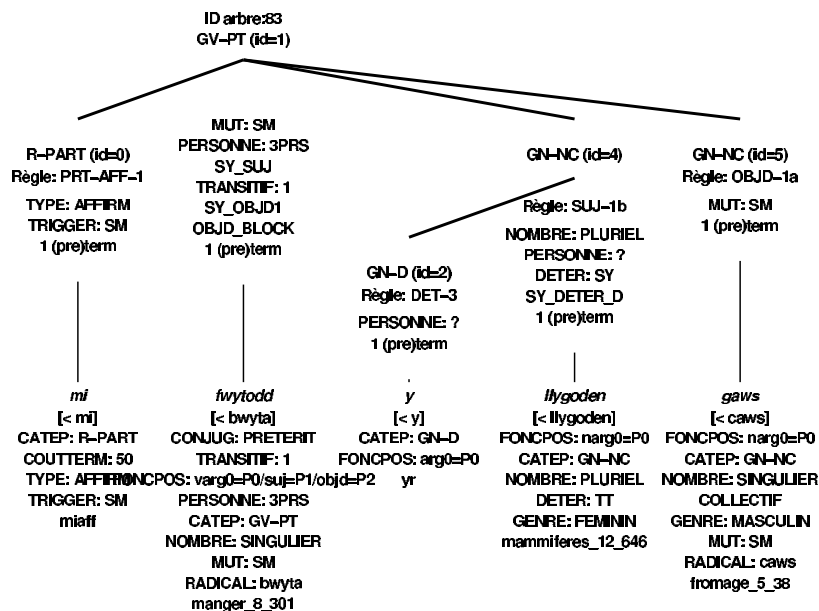


FIG. 1 – Arbre de dépendance

L’idée initiale était de vérifier si nos logiciels sont bien capables de traiter des traits typologiques du gallois absents dans les langues « bien dotées », c’est à dire la position initiale du verbe, les mutations, l’accord verbe – sujet différent selon un sujet nominal (verbe toujours en singulier) ou un sujet pronominal (accord de la personne et du nombre), ou encore la façon syntaxique de exprimer les temps sémantiques et l’aspect (*wedi, yn*). Le modèle sémantique pour le traitement du temps sémantique à été initialement présenté dans Heinecke, 1999, sa mise en œuvre est décrit dans Heinecke, 2005. Le lexique n’est pas (encore) lié à notre thésaurus sémantique ; ceci serait nécessaire pour aborder la traduction automatique dans le cadre de l’approche choisi (base sur des idées de la Discourse Representation Theory, Kamp & Reyle, 1993) en passant par un pivot sémantique, qui est indépendant de la langue source (Heinecke & Toumani, 2003). Après l’analyse sémantique on obtient un graphe sémantique (figure 2).

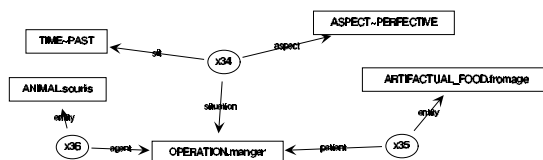


FIG. 2 – Graphe sémantique correspondant à figure 1

Malgré la structure syntaxique très différente d'une phrase qui utilise les moyens syntaxique afin d'exprimer le présent antérieur (*mae'r llygoden wedi bwyta'r caws* « la souris vient de manger le fromage ») on obtient un graphe sémantique assez similaire, à part des prédicats du temps sémantique (figure 3).

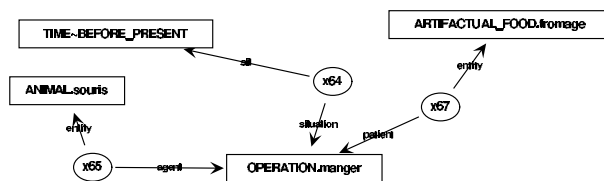


FIG. 3 – Graphe sémantique de la phrase *mae'r llygoden wedi bwyta'r caws*

Les travaux entamés seront poursuivis afin d'arriver à une traduction automatique plus complète pour au moins un domaine partiel. Le résumé automatique est l'autre objectif en vue pour notre boîte à outils linguistiques pour le gallois. Il est cependant trop tôt pour présenter des résultats.

Les expérimentations montrent que les traits typologiques du lexique et de la syntaxe galloise ne posent aucune difficulté à nos outils du TALN. Les mutations sont déjà gérées au niveau lexical. La théorie de dépendance sur laquelle notre analyseur s'appuie est assez indépendante pour permettre un ordre de mot différent du « Standard Average European », SAE.

4 Conclusion

Le *Welsh Language Board* a récemment publié une stratégie afin d'améliorer la technologie de l'information du gallois ([Welsh Language Board, 2004](#)) ainsi qu'un rapport qui propose les travaux nécessaires pour le développement d'un système de traduction automatique gallois et anglais ([Somers, 2004](#)). À côté de la recherche scientifique, il existe un besoin d'avancement des technologies du TALN et des ressources linguistiques pour le gallois, surtout dans les domaines du traitement de l'oral (reconnaissance et synthèse vocales) et dans les technologies support pour les personnes qui utilisent le gallois dans un contexte d'informatisation (par exemple correcteurs d'orthographe, traduction automatique). Nous avons montré que le gallois n'est plus une langue « peu dotée » pour la technologie de l'information, mais le nombre de ressources reste limité. Par contre on trouve des travaux actuels dans tous les domaines autour de la technologie de l'information. En ce qui concerne le TALN, les études linguistiques du gallois ont résulté en une grande diversité des descriptions linguistiques détaillées sur tous les niveaux : phonétique/phonologie, lexique ou encore syntaxe. L'existence de lexiques bilingues gallois/anglais (sur support électronique²⁴) pourra faciliter la création des données sémantiques à partir des données similaires pour l'anglais.

²⁴Malgré le fait que les données ne sont pas encore accessibles sous forme d'une base de données, le GPC ([Thomas & Bevan, 1950-2002](#)) et le dictionnaire anglais-gallois ([Griffiths & Jones, 1995](#)) sont sur support électronique et accessibles en-ligne (voir note en base de page no. 18).

Abbreviations

AFF	affirmatif	PRED	particule prédicative
AM	aspiration	PRÉT	préterit
+AM	cause l'AM sur le mot suivant	SM	lénition
IMPF	imperfectif	+SM	cause la SM sur le mot suivant
NM	nasalisation	VN	nom verbal
+NM	cause la NM sur le mot suivant		
PRÉS	présent		

Références

- Awbery G. M. (1976). *The Syntax of Welsh. A transformational Study of The Passive*. Cambridge : Cambridge University Press.
- Ball M. J. & Fife J. (2002). *The Celtic Languages*. London : Routledge.
- Ball M. J., Griffiths T. & Jones G. E. (1988). Broadcast Welsh. In M. J. Ball, Ed., *The Use of the Welsh*, p. 182–191. Clevedon : Multilingual Matters.
- Ball M. J. & Müller N. (1992). *Mutations in Welsh*. London : Routledge.
- Borsley R. D. (1990). Welsh Passives. In M. J. Ball, J. Fife & E. Poppe, Eds., *Celtic Linguistics. Festschrift for T. Arwyn Watkins*, Amsterdam studies in the theory and history of linguistic science : Series 4 ; 68. Amsterdam : John Benjamins.
- Borsley R. D. (1999). Mutation and constituent structure in Welsh. *Lingua*, **109**, 267–300.
- Ellis N. C., O'Dochartaigh C., Hicks W., Morgan M. & Laporte N. (2001). Cronfa Electroneg o Gymraeg (CEG). http://www.bangor.ac.uk/ar/cb/ceg/ceg_cym.html.
- Fynes-Clinton O. (1913). *The Welsh Vocabulary of the Bangor District*. Oxford : Oxford University Press.
- Gensler O. D. (1993). *A typological evaluation of Celtic/Hamito-Semitic syntactic parallels*. PhD thesis, University of California.
- Griffiths B. & Jones D. G. (1995). *The Welsh Academy English-Welsh Dictionary*. Cardiff : University of Wales Press.
- Heinecke J. (1999). *Temporal Deixis in Welsh and Breton*. Anglistische Forschungen 272. Heidelberg : Winter.
- Heinecke J. (2005). Temps sémantique et aspect dans l'analyse sémantique automatique. In *Actes de la plate-forme AFIA 2005. Atelier "Raisonnement du temps et espace"*.
- Heinecke J. & Toumani F. (2003). A Natural Language Mediation System for E-Commerce applications. An ontology-based approach. In ISWC, Ed., *Proceedings of Workshop Human Language Technology for the Semantic Web and Web Services. International Semantic Web Conference, Sanibel Island, Florida, 20-23 October 2003*, p. 39–50.
- Humphreys H. L. (1980). *La Langue Galloise. 2 vols. Une Présentation*. Brest : Université de Bretagne Occidentale.

- Jones D. G. (1988). Literary Welsh. In M. J. Ball, Ed., *The Use of the Welsh*, p. 125–171. Clevedon : Multilingual Matters.
- Jones R. J., Mason J. S., Jones R. O., Helliker L. & Pawlewski M. (1998). SpeechDat Cymru. A Large-scale Welsh Telephony Database. In *Language Resources for European Minority Languages, 1998. Proceedings of the LREC Workshop, Granada, Spain 1998*.
- Kamp H. & Reyle U. (1993). *From Discourse to Logic. Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Studies in Linguistics and Philosophy 42. Dordrecht : Kluwer.
- King G. (1993). *Modern Welsh. A comprehensive grammar*. London, New York : Routledge.
- Morris-Jones J. (1913). *A Welsh Grammar. Historical and Comparative*. Oxford : Clarendon Press.
- Morris-Jones J. (1921). *An Elementary Welsh Grammar*. Oxford : Clarendon Press.
- Pokorny J. (1927-30). Das nicht-indogermanische Substrat im Irischen. *Zeitschrift für Celtische Philologie*, 16 + 17.
- Roberts I. G. (2004). *Principles and Parameters in a VSO Language. A Case Study in Welsh*. Oxford Studies in Comparative Syntax. Oxford : Oxford University Press.
- Rouveret A. (1990). X-Bar Theory, Minimality, and Barrierhood in Welsh. In H. Randall, Ed., *The Syntax of the Modern Celtic Languages*, Syntax and Semantics 23, p. 27–77. New York : Academic Press.
- Sadler L. (1998). Welsh NPs without Head Movement. In M. Butt & T. H. King, Eds., *Proceedings of the LFG98 Conference*, Stanford : CSLI Publications.
- Sadler L. (1999). Non-Distributive Features in Welsh Coordination. In M. Butt & T. H. King, Eds., *Proceedings of the LFG99 Conference*. Stanford : CSLI Publications.
- Sadler L. (2003). Function Spreading in Coordinate Structures. Paper given at the 4th Celtic Linguistics Conference. <http://privatewww.essex.ac.uk/~louisa/newpapers/tense-share2.pdf>.
- Somers H. (2004). Machine Translation and Welsh. The Way Forward. <http://www.bwrdd-yr-iaith.org.uk/en/cynnwys.php?cID=6&pID=109&nID=1190>.
- Thomas A. R. (1973). *Linguistic Geography of Wales*. Cardiff : University of Wales Press.
- Thomas A. R. (1987). A spoken standard for Welsh. Description and Pedagogy. In G. Williams, Ed., *The Sociology of Welsh*, p. 99–113. Berlin : Mouton de Gruyter.
- Thomas B. & Thomas P. W. (1989). *Cymraeg, Cymrâg, Cymrêg...Cyftwyno'r Tafodieithoedd*. Caerdydd : Gwasg Tâf.
- Thomas P. W. (1996). *Gramadeg y Gymraeg*. Caerdydd : Gwasg Prifysgol Cymru.
- Thomas R. J. & Bevan G. A. (1950-2002). *Geiriadur Prifysgol Cymru. A Dictionary of the Welsh Language*. Caerdydd : Gwasg Prifysgol Cymru.
- Thorne D. (1993). *A Comprehensive Welsh Grammar*. Oxford : Blackwell.
- Vennemann T. (2002). Semitic → Celtic → English : The transitivity of language contact. In M. Filppula, J. Klemola & H. Pitkänen, Eds., *The Celtic roots of English*, Studies in Languages 37, p. 295–330. Joensuu : University of Joensuu.

Wagner H. (1959). *Das Verbum in den Sprachen der Britischen Inseln*. Tübingen : Niemeyer.

Welsh Language Board (2004). *Information Technology and the Welsh Language. A strategy document*.

Williams B. (1999). A Welsh speech database. Preliminary result. In *EuroSpeech 1999. Proceedings of the Sixth European Conference on Speech Communication and Technology, Budapest, Hungary, September 5-9, 1999*, Budapest.

Williams S. J. (1980). *Elements of a Welsh Grammar*. Cardiff : University of Wales Press.