

GROBID-DICTIONARIES: INFRASTRUCTURE FOR AUTOMATICALLY STRUCTURING DIGITIZED DICTIONARIES AND ENTRY-BASED DOCUMENTS

MOHAMED KHEMAKHEM
MOHAMED.KHEMAKHEM@INRIA.FR

Context: PhD Thesis

Standard-based Lexical Models for Automatically Structured Dictionaries

Context: PhD Thesis

Standard-based Lexical Models for Automatically Structured Dictionaries

Automatically Structuring: Landscape

- Domination of rule based methods

- Just 2 previous works used machine learning
 - Crist, S. (2011)
 - Bago, P. and Ljubešić, N. (2015)

Standard-based Lexical Models

- Standardized resources are required for large scale projects
 - Enable mapping and enriching

- Text Encoding Initiative (TEI)
 - Too flexible

- Lexical Markup Framework (ISO 24613 LMF)
 - Too restricted

GeneRation Of Bibliographic Data = GROBID



GROBID: GeneRation Of Bibliographic Data

- Java library for processing and structuring text information (mainly in PDF)
- Supervised machine learning for Information Extraction
 - Sequence Labelling: Conditional Random Fields (CRF)
- TEI-compliant output
- Used by: Research Gate, HAL, Mendeley, CERN,...

GROBID: GeneRation Of Bibliographic Data

Depth-resolved analysis of spontaneous phase separation in the growth of lattice-matched AlInN

title

A. Redondo-Cubero^{1,2,*}, K. Lorenz³, R. Gago⁴, N. Franco³, M.-A. di Forte Poisson⁵, E. Alves³ and E. Muñoz¹

authors

- 1 ISOM and Dpt. de Ingeniería Electrónica, ETSI Telecomunicación, Universidad Politécnica de Madrid, E-28040 Madrid, Spain.
- 2 Centro de Micro-Análisis de Materiales, Universidad Autónoma de Madrid, E-28049 Madrid, Spain.
- 3 Instituto Tecnológico e Nuclear, Estrada nacional 10, 2686-953 Sacavém, Portugal.
- 4 Instituto de Ciencia de Materiales de Madrid (CSIC), E-28049 Madrid, Spain.
- 5 Thales Research & Technology/TIGER, 91461 Marcoussis Cedex, France.

affiliation

ABSTRACT:

We report the detection of phase separation of an $\text{Al}_{1-x}\text{In}_x\text{N}/\text{GaN}$ heterojunction grown close to lattice matched conditions ($x \sim 0.18$) by means of Rutherford backscattering spectrometry in channeling geometry and high resolution x-ray diffraction. An initial pseudomorphic growth of the film was found, with good single crystalline quality, the

abstract

GROBID-Family

GeneRation Of Bibliographic Data = GROBID



Dictionary Segmentation Level

CON

Eugène IV à Ferrare en 1438-1439, puis à Florence de 1439 à 1442, de Laurent (1417-1517), de Trente (1545-1563) (où fut décidée son réforme générale de l'Église catholique en face de la Réforme protestante), de Vatican I (1870) (où fut défini le dogme de l'Immaculée Conception), de Vatican II (1962-1965) (où fut définie l'attitude de l'Église romaine à l'égard du monde moderne).

conciliable adj. Qui peut se concilier avec une autre chose.

conciabulaire [kɔ̃siablɔʁ] n. m. (lat. *conciabularium*). Réunion secrète de personnes soupçonnées de mauvais desseins : *tenir des conciabulaires*. Entretien plus ou moins secret et suspect.

conciataire adj. Relatif à un concile : *décret conciataire*.

conciant, e adj. Porté à la conciliation : *conciante conciliation*. | Progre à concilier : *des paroles conciliantes*.

conciataire, trice n. Personne qui concilie, sime à concilier.

conciliation n. f. Action de concilier ; résultat de cette action. | Accord de deux personnes en litige, réalisé par un juge.

conciliataire adj. Propre à concilier : *mœurs conciliatoires*.

concilier [kɔ̃silje] v. t. (lat. *conciliare*). Mettre d'accord ; concilier des adversaires.

— Se concilier v. pr. Acquiescer, s'entier : se concilier l'estime d'autrui.

conclis [kɔ̃sli], e [sli] adj. (lat. *conclusus*, tranché). Qui exprime beaucoup de choses en peu de mots. (Syn. : bref, court, net, précis, succinct).

conclusion n. f. Qualité de ce qui est conclis.

concluyen, enne n. Qui est au même pays, de la même ville.

conclavé [kɔ̃klav] n. m. (lat. *conclave*). Assemblée de cardinaux pour élire un pape.

conclaviste n. m. Personne qui s'entretient au conclavé avec un cardinal, pour le servir.

concluant, e adj. Qui prouve ce qu'on a avancé : *argument concluant*.

conclure [kɔ̃kyʁ] v. t. (lat. *concludere*) [concl. 42]. Achever, terminer : *conclure une affaire*. Triser une conversation : *conclure une chose d'une suite*. — V. Donner son avis, ses conclusions ; se promettre : *on s'est demandé de conclure*. Être proche des témoignages concluent contre lui.

conclusion n. f. (lat. *conclusio*). Arrangement défini ; *conclusion d'un traité* ; *Fin, résultat final* ; la conclusion d'un argument d'un syllogisme ne doit pas être la conclusion d'un syllogisme ne doit pas être la conclusion d'un syllogisme.

— Pré. Prétention respectives de chacune des parties dans un procès ; *Écrit expojant ces prétentions*. | Réquisition du ministre public.

— En conclusion loc. adv. En conséquence, pour conclusion.

concocter v. t. Fam. Elaborer avec soin ; concocter une lettre de réclamation.

concombre [kɔ̃kɔ̃br] n. m. (anc. provenç. *comcombre*). Plante potagère de la famille des cucurbitacées, cultivée pour ses fruits allongés que l'on consomme comme légume ou en salade. | Ce fruit.

concomitamment adv. De façon concomitante.

concomitance [kɔ̃kɔ̃mitɑ̃] n. f. Coexistence, simultanéité de deux ou plusieurs faits.

concomitant, e adj. (lat. *concomitans*). Qui accompagne, qui se produit en même temps : *des faits concomitants*. ■ *Épithètes concomitantes*, variations simultanées et proportionnelles de certains phénotypes.

concordance n. f. Conformité, accord ; concordance de témoignages. | *Général*. Disposition parallèle des cochets sédimentaires. ■ *Concordance de phases* (Phys.), état de plusieurs vibrations sinusoïdales de même nature et de même période, dont la différence de phases est nulle. *Concordance des tempétries* de syntaxe d'après lesquelles le temps du verbe d'une subordonnée varie selon celui du verbe de la principale.

concordant, e adj. Qui s'accorde : *énoncés concordants* (1847).

concordat [kɔ̃kɔ̃rdɑ̃] n. m. (lat. *concordatum*). Traité entre le pape et un gouvernement sur les affaires religieuses.

concorder v. t. (lat. *concordare*). Faire le commerçant qui, ayant déposé son bilan, a été admis par le tribunal de commerce au règlement judiciaire et ses créanciers.

Les plus anciens concordats sont le concordat de Worms (1122), entre Calixte II et Henri V.

Le concordat de 1516, entre François I^{er} et Léon X et François I^{er}. Le concordat entre Bonaparte et Pie VII.

Le concordat du 16 juillet 1801, a réglé les rapports de la France avec le Saint-Siège, et est signé avec l'Église jusqu'à la loi du 9 décembre 1905. Au XIX^e s. et au XX^e s., de nombreux concordats furent signés par les papes.

concordiaire adj. Relatif à un concordat : *loi concordiaire*. *Dr.* Se dit du commerçant qui a obtenu un concordat.

concorde n. f. (lat. *concordia*). Accord des sentiments et des volontés : *rétablir la concorde entre citoyens*.

concorde [kɔ̃kɔ̃rdɑ̃] v. t. (lat. *concordare*). Avoir des rapports de similitude, de correspondance : *dates qui concordent*.

concordeur, e adj. Qui converge vers un même point, un même but : *droites concordeuses*.

concorreur [ɑ̃] v. t. ind. (lat. *concurrere*) [concl. 21]. Tendre au même but, aider à : *concorreur au succès d'une affaire*. — V. Être en concurrence, en compétition : *concorreur pour un prix*.

concoeurs [kɔ̃kyʁ] n. m. (lat. *concoeurs*). Concurrence, concurrence ; *concoeurs de consciences*. | Action de concourir, d'aider : *concoeurs pour un prix*.

concoeurs [kɔ̃kyʁ] n. m. (lat. *concoeurs*). Concurrence avec d'autres, pour prétendre à quelque chose ; exaimes : *se présenter à un concoeurs*. | Lutte sportive : *concoeurs hipique*.

■ *Concoeurs général*, concoeurs annuel entre les premiers élèves des classes supérieures des lycées, collèges et écoles normales.

concoeur [kɔ̃kyʁ] n. m. (lat. *concoeur*). Épave, condamné ; *huile concoeur* (huile).

concoeur [kɔ̃kyʁ] n. m. (lat. *concoeur*). Qui se présente comme chose de justice, de positif ; *obtenir des avantages concoeurs*. | Qui le fait des idées précises des esprits concoeurs. | *Gramm.* Se dit d'un terme qui désigne, et qui peut toujours être perçu par les sens. ■ *Musique concoeur*. Technique de composition qui utilise les bruits produits par divers objets sonores enregistrés sur bande magnétique et susceptibles de transformation.

concret n. m. Qualité de ce qui est concret.

concrètement adv. De façon concrète.

concréter v. t. (concl. 5). Rendre concret, solidifier.

concrétion [kɔ̃kɔ̃ʁeʃ] n. f. (cde *concre*). Action de concrétiser ; la concrétion de l'urine.

concréter v. t. (concl. 5). Rendre concret ce qui est abstrait : *concréter une idée, un avantage*.

concrubilité n. f. (lat. *concrubilis*). Relatif au concrubage. — N. Personne qui vit en concrubage.

concrubage [kɔ̃kɔ̃ʁabʒ] n. m. m. Etat d'un individu et d'une femme qui vivent ensemble sans être mariés. (On dit aussi union libre).

concupiscence n. f. (du lat. *concupiscere*, désirer). Penchant à jouir des biens terrestres, particulièrement des plaisirs sensuels.

concupiscence [kɔ̃kypɛ̃sɑ̃] n. f. (-at) adj. Qui a été adonné par le tribunal de commerce au règlement judiciaire et ses créanciers.

concurrance [kɔ̃kyʁɑ̃sɑ̃] adv. Par concurrence. | Par un concours mutuel, de concert avec : *concurrance avec quelqu'un*.

concurrance n. f. Rivalité entre plusieurs personnes qui visent un même but ; *entrer en concurrance avec quelqu'un*. | Rivalité d'intérêts entre commerçants ou industriels qui tentent d'attirer à eux la clientèle par les meilleures conditions de prix, de qualité, etc. ■ *Régime de libre concurrance*, système économique qui ne comporte aucune intervention de l'État en vue de limiter la liberté de l'industrie et de commerce, et qui considère les coalitions de producteurs comme des délits. — Jusque à *concurrance* de loc. prép. Jusqu'à la somme de.

concurrer v. t. (concl. 1). Faire concurrence à.

concurrer [kɔ̃kyʁɑ̃] e [-at] adj. et n. Qui tend au même but ; *une action concurrer*. | Personne qui participe à un concours, à une compétition : *les concurrents ont pris le départ de la course*. | Celui qui entre la même profession commerciale qu'un autre.

concurreriel [kɔ̃kyʁɑ̃siel] n. m. et n. Qui loue la concurrence.

concurreriel [kɔ̃kyʁɑ̃siel] n. f. (lat. *concurreriel*). Exaction commise par un trésorier public, l'administration communale l'exercice d'une fonction publique, particulièrement dans le maniement des deniers publics.

concurreriel [kɔ̃kyʁɑ̃siel] n. m. et n. Coupable de concurrence.

condamnabile adj. Qui mérite d'être condamné : *acte condamnable*.

condamnation [kɔ̃damɑ̃sɑ̃] n. f. (lat. *condamnation*). Jugement prononcé par un tribunal, imposant à l'un des plaideurs de s'incliner au moins partiellement devant les prétentions de son adversaire. | Décision d'une juridiction prononçant contre une personne d'un crime, d'un délit ou d'une contravention. (En matière de droit pénal, on distingue de l'accusé, et la cour prononce la condamnation. La condamnation est la sanction infligée à la *déclaration criminelle*. ■ *Bilème*, désapprobation ; la *condamnation des autres*.)

condamner v. t. (concl. 5). Faire porter condamnation.

condamné, e n. Personne qui a subi une condamnation. — Adj. Qui ne peut échapper à un sort prévu : *malade condamné*.

condamner [kɔ̃damɑ̃] v. t. (lat. *condamnation*). Prononcer un jugement de plaideur ou un inculpé : *condamner un criminel*. | *Astérisation*. | *Agitation* : *condamner à l'immobilité*, à l'indépendance, à l'immobilité, à l'indépendance, à l'immobilité.

condamner [kɔ̃damɑ̃] v. t. (lat. *condamnation*). Interdire : *la loi condamne la bigamie*. | Déclarer perdu, incurable : *les médecins l'ont condamné*. | Barrez, courez ; *condamner une porte*.

condamnable adj. Qui peut être condamné, réduit à un moindre volume.

condamner n. m. Appareil servant à emmagasiner une charge électrique ; *une batterie de Leyde* est un condamnateur électrique. | Lentille servant à éclairer un objet ; *une lampe à incandescence*.

condensation n. f. Action de condenser et effet qui en résulte ; liquéfaction d'un gaz : *Soudure de plusieurs métaux chimiques, avec élimination d'eau*.

condensation n. m. Résumé d'une œuvre littéraire.

condenser [kɔ̃dɑ̃sɑ̃] v. t. (lat. *condensare*, rendre épais). Rendre plus dense, réduire à un moindre volume. | Qualifier un gaz par refroidissement ou compression : *le froid condense la vapeur*. | *Fig.* Exprimer d'une manière concise, en peu de mots.

Dictionary Body Segmentation Level

CON

Enguine IV à Ferrate en 1438-1439, puis à Florence de 1439 (1442), à La Rochelle (1517), de Tienne (1543-1563) [ou fut décidée la réforme générale de l'Église catholique en face de la Réforme protestante], de Vailly (1870) [ou fut défini le dogme de l'Infaillibilité pontificale], de Vailly II (1962-1965) [ou fut définie l'attitude de l'Église romaine à l'égard du monde moderne].

conciliabule [kɔ̃siljabyl]. n. m. (lat. conciliabulum). Réunion secrète de personnes soupçonnées de mauvais desseins : *tenir des conciliabules*. | Entrées plus ou moins secrètes et suspectes.

conciliabile adj. Relatif à un concile. *décider conciliabile*.

concilier [kɔ̃silje]. v. t. (lat. conciliare). Faire concilier. | Propre à concilier : *une attitude conciliante*.

conciliateur, trice n. Personne qui concilie, aime à concilier.

conciliation n. f. Action de concilier ; résultat de cette action. | Accord de deux personnes en litige, traité, etc. un *juste conciliateur* adj. Propre à concilier : *meures conciliateures*.

concilier [kɔ̃silje]. v. l. (lat. conciliare). Mettre d'accord : *concilier des adversaires*.

— **Le concilier** v. pr. Acquisir, s'attirer : *se concilier l'estime d'autrui*.

concia [kɔ̃si]. n. [2] adj. (lat. concia). Tranche, Qui exprime beaucoup de choses en peu de mots. (Syn. : bref, court, dense, facile, succinct).

concia n. f. Qualité de ce qui est concis.

conciopon [kɔ̃siɔpɔ̃]. n. Qui est du même pays. de la même ville.

conclave [kɔ̃klav]. n. m. (lat. concilium). Assemblée de cardinaux pour élire un pape.

conclaviste n. m. Personne qui s'enferme au conclave avec un cardinal, pour le servir.

concluant e adj. Qui prouve bien ce qu'on a avancé : *argument concluant*.

conclure [kɔ̃klyʁ]. v. t. (lat. concludere) [com. 42]. Achèver, terminer : *conclure une affaire*. | Tirer une conséquence : *conclure que chose d'une autre*. | Donner son avis, ses conclusions, se prononcer : *un avis concluant*. | Être probant : *les témoignages concluent contre lui*.

conclusion n. f. (lat. conclusio). Arrangement définitif : *conclusion d'un traité*. | Fin du traité final : *conclusion d'un discours*. | Conclusion d'un argument : la *conclusion d'un syllogisme ne doit pas dépasser les prémisses*. | — Pl. Prétentions respectives de chacune des parties dans un procès. | Écrit exposant ces prétentions. | Réquisition du ministère public. | — En conclusion loc. adv. En conclusion, puis.

concocter v. l. Fam. Elaborer avec soin ; concocter une lettre de réclamation.

concombre [kɔ̃kɔ̃br]. n. m. (anc. provenç. comombro). Plante fourrue de la famille des cucurbitacées, cultivée pour ses fruits allongés que l'on consomme comme légumes ou en salade. | Ce fruit.

concombrément adv. De façon concomitante.

concomitance [kɔ̃kɔ̃mitãs]. n. f. Coexistence, simultanéité de deux ou de plusieurs faits.

concomitant e adj. (lat. concomitans). Qui accompagne, qui se produit en même temps : *des faits concomitants*. ■ Variations concomitantes, variations simultanées et proportionnelles de certaines pleurocoudes.

concordance n. f. Conformité, accord : *concordance de témoignages*. | Gr. Disposition parallèle des coécumes sédimentaires. ■ Concordance de phases (Ph. s.). État de plusieurs vibrations sinusoidales de même nature et de même période, dont la différence de phases est nulle. | Concordance des températures de synthèse d'avants lesquelles le temps du verbe d'une subordonnée varie selon celui du verbe de la principale.

concordant e adj. Qui s'accorde : *témoin concordant*.

concordat [kɔ̃kɔ̃dat]. n. m. (lat. concordatum). Traité entre le pape et un gouvernement sur les affaires religieuses. | Accord entre le comarçant qui, ayant déposé son bilan, a été admis par le tribunal de commerce au règlement judiciaire et ses créanciers.

Les plus anciens concordats sont le concordat de Worms (1122), entre Caliste II et Henri V. Le concordat de 1516, entre Léon X et François I^{er}. Le concordat entre le Saint-Siège et l'Espagne (l'Église jusqu'à la loi du 9 décembre 1965. Au XIX^e s. et au XX^e s. de nombreux concordats furent signés par les papes.

concordatoire adj. Relatif à un concordat : *loi concordatoire*. | Dr. Se dit du comarçant qui a obtenu un concordat.

concordé n. f. (lat. concordat). Accord des sentiments et des volontés : *rétablir le concordé entre les citoyens*.

concordeur [kɔ̃kɔ̃deʁ]. v. l. (lat. concordare). Avec des rapports de similitude, de correspondance : *dates qui concordent*.

concordeur e adj. Qui converge vers un même point, un même but : *droites concordeures*.

concouir [kɔ̃kuwa]. v. t. (lat. concurrere) [com. 21]. Tendre au même but, aider à : *concouir au succès d'une affaire*. | — V. l. Être en concurrence, en compétition : *concouir pour un prix*.

concouir [kɔ̃kuwa]. v. m. (lat. concurrere). Coïncidence, conjonction : *concouir de circonstances*. | Action de concourir, d'offrir son concours. | Action d'entrer en concurrence avec d'autres, pour obtenir à quelque chose ; examen : se *présenter à un concours*. | Lutte sportive : *concours hipérique*. ■ *Concours général*, concours annuel entre les premiers élèves de classes supérieures des lycées, collèges et écoles normales.

concret [kɔ̃kʁɛ]. n. m. [2] adj. (lat. concretus). Qui exprime quelque chose de réel, de positif : *obtenir des avantages concrets*. | Qui a le sens des réalités précises : *esprit concret*. | Gram. Qui est, d'un terme qui désigne un être ou un objet pouvant être perçu par les sens. ■ *Moules concrets*, technique de composition qui utilise les bruits produits par divers objets sonores enregistrés sur bande magnétique et susceptibles de transmission.

concret n. m. Qualité de ce qui est concret.

concreter [kɔ̃kʁɛ]. v. t. (lat. concretare) [com. 3]. Remettre concret, rendre concret.

concrétion [kɔ̃kʁɛsjɔ̃]. n. f. (de concrere). Solide : *concrétion osseuse*. | Aggrégation fluide dans les tissus vivants : *concrétion biliaire*.

concrétier v. l. Rendre concret ce qui est abstrait : *concrétiser une idée, son avantage*.

concupiscence n. f. (lat. concupiscentia). Relatif au concubinage. | — N. Personne qui vit en concubinage.

concupiscence [kɔ̃kypisãs]. n. f. État d'un homme et d'une femme qui vivent ensemble sans être mariés. (On dit aussi union libre.)

concupiscence n. f. (du lat. concupiscere, désirer). Penchant à pour des biens terrestres, particulièrement des plaisirs sensuels.

concupiscence [kɔ̃kypisãs]. n. f. [1-2] adj. Qui exprime la concupiscence ; regardé *concupiscence*. | Attaché aux plaisirs sensuels.

concurrent [kɔ̃kypã]. n. m. (lat. concurrere). Qui concourt avec quelqu'un. | Réalité d'indépendance entre commerçants ou industriels qui tentent d'attirer à eux la clientèle sur les meilleures conditions de prix, de qualité, etc. ■ *Régime de libre concurrence* système économique qui ne comporte aucune intervention de l'État en vue de limiter la liberté de l'industrie et du commerce, et qui considère les coalitions de producteurs comme des délits. | — Jusqu'à *assurance de la vie* par *jusqu'à la tombe*.

concurrent v. l. (lat. concurrere). 1) Faire concurrence à.

concurrent [kɔ̃kypã]. n. m. [1-2] adj. et n. Qui tend au même but, a une action concurrente. | Personne qui participe à un concours, à une compétition ; les *concurrents ont pris le départ de la course*. | Celui qui exerce la même profession commerciale qu'un autre.

concurrentiel [kɔ̃kypãsjɔ̃]. n. m. adj. Ou l'ou la concurrence.

concurrentiel [kɔ̃kypãsjɔ̃]. n. f. (lat. concurrere). Exécution commise par un rétrorier public. | Malversation commise dans l'exercice d'une fonction publique, particulièrement dans le manement des deniers publics.

concessionnaire adj. et n. Coupable de concession.

condamnabile adj. Qui mérite d'être condamné.

condamnation [kɔ̃damãs]. n. f. (lat. damnatio). Déclat d'un tribunal imposé à l'un des plaignants de l'acte incriminé, moins particulièrement devant les prétentions de son adversaire. | Déclat d'une juridiction prononçant une peine contre l'auteur d'un crime, d'un délit ou d'une contravention (en cour d'assises, le jury juge la culpabilité par l'accusé et la cour prononce la condamnation). | La peine infligée ; une *condamnation à la réclusion criminelle*. | Blâme, désapprobation : la *condamnation des idées*.

condamner [kɔ̃damã]. v. t. Qui porte condamnation.

condanné e n. Personne qui a subi une condamnation. | — Adj. Qui ne peut échapper à une condamnation.

condanné [kɔ̃damã]. v. l. (lat. condemnare). Prononcer la condamnation de quelqu'un ou d'un inculpé ; *condanné un criminel*. | Astronomie, réduire à l'équateur : *condanné la lune*, à l'immobilité. | Désapprouver, blâmer : *condanné une opinion, un usage*. | Interdire, le loi condonne la bigamie. | Déclarer perdu, incurable : *le médecin l'a condanné*. | Biter, couter ; *condanné une porte*.

condannable adj. Qui peut être condanné, réduit à un moindre volume.

condanné [kɔ̃damã]. v. m. [1-2]. Appareil servant à emmagasiner une charge électrique : *condanné de Leyde* est un condensateur électrique. | Lentille servant à éclairer un objet qui doit former une image.

condanné n. f. ACTION DE CONDANNER ou effet qui en résulte. | L'opération qui consiste à souder de plusieurs molécules chimiques, avec élimination d'eau.

condanné n. m. Résumé d'une œuvre littéraire.

condanné [kɔ̃damã]. v. l. (lat. condensare). rendre épais. | Rendre plus dense, réduire à un moindre volume. | Liquefier un gaz par refroidissement ou compression : *le froid condanné la vapeur d'eau*. Fig. expression d'une manière concise : *un mot de moins*.

Lexical Entry Leve

CON

Eugène IV > Ferrare en 1438-1439, puis à Florence de 1459 à 1462, de Lausanne (1512-1517), de Trente (1543-1563) (où fut décidée la réforme générale de l'Église catholique en face de la Réforme protestante), de Vatican I (1870) (où fut défini le dogme de l'Immaculée Conception), de Vatican II (1962-1965) (où fut définie l'attitude de l'Église romaine à l'égard du monde moderne).

conciliable adj. Qui peut se concilier avec une autre chose.

concubinaire [kɔ̃siljabɛ̃] n. m. (lat. *concupubium*). Réunion secrète de personnes soupçonnées de mauvais desseins : *tenir des concubinaires*. | Entrées plus ou moins secrètes et suspectes.

conciliant adj. Relatif à un concile : *décret conciliant*.

conciliant e adj. Prové à la conciliation : *caractères conciliants*. | Propre à concilier : *des paroles conciliantes*.

conciliateur, trice n. Personne qui concilie, aime à concilier.

conciliation n. f. Action de concilier ; résultat de cette action. | Accord de deux personnes en litige, traité par un juge.

conciliateur adj. Propre à concilier : *mœurs conciliateuses*.

concilier [kɔ̃silje] v. t. (lat. *conciliare*). Mettre d'accord : *concilier des adversaires*. | — Se concilier v. pr. Acquiescer, s'attirer : *se concilier l'estime d'autrui*.

concordant e adj. Qui s'accorde : *témoin concordant*.

concordat [kɔ̃kɔʁda] n. m. (lat. *concordatum*). Traité entre le pape et un gouvernement sur les affaires religieuses. | Dr. Accord entre le commerçant qui, ayant déposé son bilan, a été admis par le tribunal de commerce au règlement judiciaire et ses créanciers.

Les plus anciens concordats sont le concordat de Worms (1122), entre Calixte II et Henri V, le concordat de 1516, entre Léon X et François I^{er}. Le concordat entre le Saint-Siège, et l'empereur d'Autriche jusqu'à la loi du 9 décembre 1965. Au XIX^e s. et au XX^e s. de nombreux concordats furent signés par les papes.

concordataire adj. Relatif à un concordat : *loi concordataire*. | Dr. Se dit du commerçant qui a obtenu un concordat.

concordé n. f. (lat. *concordia*). Accord des sentiments et des volontés : *rétablir la concordé entre les citoyens*.

concordeur [kɔ̃kɔʁde] v. t. (lat. *concordare*). Avec des rapports de similitude, de correspondance : *dates qui concordecnt*.

concordeur e adj. Qui converge vers un même point, un même but : *droites concordecntes*.

concouvri v. t. (lat. *concupravit*) [kɔ̃ki]. Tendre au même but, aider à :

concupiscence n. f. (du lat. *concupiscere*, désirer). Penchant à pour des biens terrestres, particulièrement des plaisirs sensuels.

concupiscent [kɔ̃ki] adj. Qui éprouve la concupiscence ; regardé *concupiscent*. | Attaché aux plaisirs sensuels.

concurrent [kɔ̃kɔʁdɑ̃] n. m. (lat. *concurrere*). Par un concours mutuel, de concert : *agir concurrentement avec quelqu'un*.

concurrance n. f. Rivalité entre plusieurs personnes qui visent un même but : *entrer en concurrance avec un concurrent*. | Réalité d'indépendants entre commerçants ou industriels qui tentent d'attirer à eux la clientèle par les meilleures conditions de prix, de qualité, etc. | Régime de libre concurrence système économique qui ne comporte aucune intervention de l'État en vue de limiter la liberté de l'industrie et du commerce, et qui considère les coalitions de producteurs comme des délits. | Jusqu'à assurance de la vie par jusqu'à la tombe de :

concurrance v. t. (lat. *concurrere*) [kɔ̃ki].

concurrer [kɔ̃ki] v. t. (lat. *concurrere*) [kɔ̃ki].

concurrer e adj. Qui tend au même but : *un concurrent*. | Personne qui participe à un concours, à une compétition ; *les concurrents ont pris le départ de la course*. | Celui qui exerce la même profession commerciale qu'un autre.

condenser [kɔ̃dɑ̃sɛ] v. t. (lat. *condensare*, rendre épais). Rendre plus dense, réduire à un moindre volume. | Liquéfier un gaz par refroidissement ou compression : *le froid condense la vapeur d'eau*. | Fig. Exprimer d'une manière concise, en peu de mots :

concret n. m. Qualité de ce qui est concret.

concréter [kɔ̃kʁete] v. t. (lat. *concretere*) [kɔ̃kʁete].

concréter v. t. (lat. *concretere*) [kɔ̃kʁete].

concrétion [kɔ̃kʁete] n. f. (de *concretere*). Action de s'épaissir : *la concrétion de l'urine*.

concrétion [kɔ̃kʁete] n. f. (de *concretere*). Réunion de parties en un corps solide : *concrétion osseuse*. | Agrégation solide dans les tissus vivants : *concrétions biliaires*.

concrétier v. t. Rendre concret ce qui est abstrait : *concréter une idée, un avantage*.

concupine e adj. (lat. *concupine*). Relatif au concubinage. | — N. Personne qui vit en concubinage.

concubinage [kɔ̃kɔʁbi] n. m. État d'un homme et d'une femme qui vivent ensemble sans être mariés. (On dit aussi *union libre*.)



condanné e n. Personne qui a subi une condamnation. | — Adj. Qui ne peut échapper à un sort néfaste : *malade condamné*.

condamner [kɔ̃dɑ̃ne] v. t. (lat. *condemnare*). Prononcer une sentence contre quelqu'un plus dur ou un coupable : *condamner un criminel*. | Astronomie, réduire à : *condamner au silence, à l'immobilité*. | Désapprouver, blâmer : *condamner une opinion, un usage*. | Interdire : *la loi condamne la bigamie*. | Déclarer perdu, incurable : *la médecine l'ont condamné*. | Biter, mater ; *condamner une porte*.

condensable adj. Qui peut être condensé, réduit à un moindre volume.

condensateur n. m. *Phys.* Appareil servant à emmagasiner une charge électrique. | *Automobile*. Le condensateur est un condensateur électrique. | Lentille servant à élargir un faisceau lumineux.

condensation n. f. Action de condenser ou effet qui en résulte. | *Phys.* État d'un gaz refroidissement ou compression ; *le froid condense la vapeur d'eau*. | Fig. Exprimer d'une manière concise, en peu de mots :

GROBID-Dictionaries: Architecture



CRF model

Dictionary Segmentation

Segmented Page

headnote

body

footnote

<dictScrap>

CRF model

Dictionary Body Segmentation

Segmented Page

<entry>

<dictScrap>

CRF model

Lexical Entry

Segmented LE

<form>

<etym>

<sense>

<re>

<dictScrap>

CRF models

Form

Sense

Segmented Form/Sense

<orth>

<pron>

<gramGrp>

<gramGrp>

<sense>

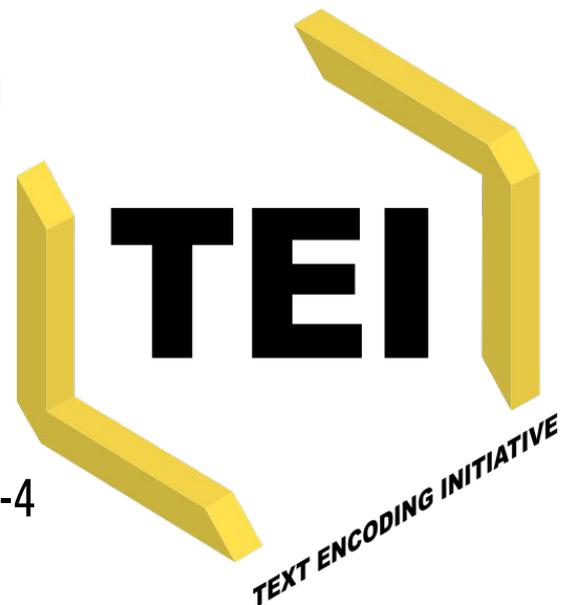
<dictScrap>

CRF model

Grammatical Group

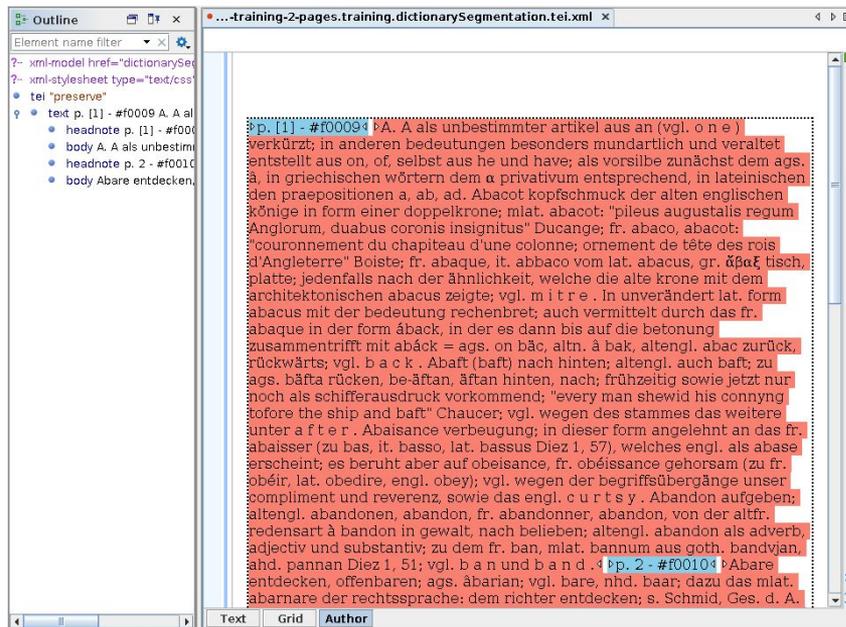
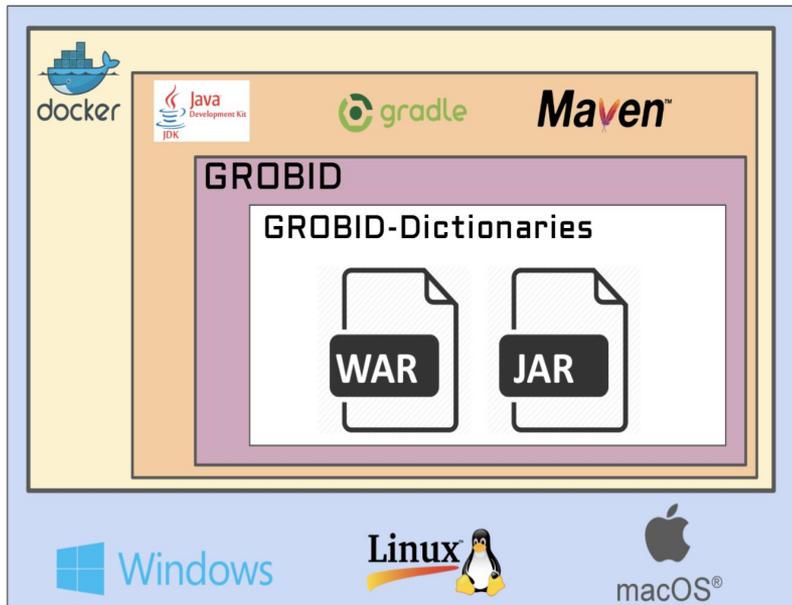
GROBID-Dictionaries: Output Serialisation

- TEI-Lex-0
 - New recommendations for unified TEI modelling
- LMF
 - Review the meta model
 - Propose a new TEI-based serialisation: ISO 24613-4



GROBID-Dictionaries: Enhanced Usability

- Khemakhem et al.2018a



Use cases: Lexical Dictionaries

- Khemakhem et al. 2017
 - Monolingual digital English dictionary
 - Bilingual digitised Fang-French dictionary

```

===== Token-level results =====

```

label	accuracy	precision	recall	f1
<form>	99.69	99.31	97.28	98.28
<pc>	99.39	100	78.79	88.14
<re>	90.49	0	0	0
<sense>	88.57	85.18	98.64	91.42

Evaluation of "Lexical Entry" model on EEBD

```

===== Token-level results =====

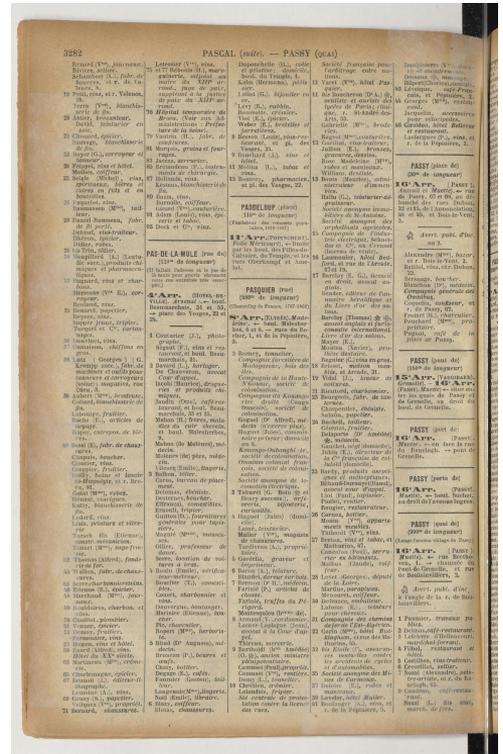
```

label	accuracy	precision	recall	f1
<form>	90.77	57.94	75.26	65.47
<pc>	97.6	28.57	11.76	16.67
<sense>	92.45	96.46	93.59	95

Evaluation of "Lexical Entry" model on FFD

Use cases: Legacy Address Directories

- Khemakhem et al. 2018c



PAS-DE-LA-MULE (rue du)
(120^m de longueur)

(Il fallait l'adresse et le pas de la mule pour graver sûrement cette rue autrefois très escarpée.)

4^e Arr. (HOTEL-DE-VILLE). Arsenal. ← boul. Beaumarchais, 31 et 33, → place des Vosges, 22 et 24.

1 Couturier (J.), *photographe.*

Ségaut (F.), *vins et restaurant,* et boul. Beaumarchais, 31.

2 Bavard (L.), *horloger.*
De Chauveron, *avocat Cour d'appel.*
Iacobi (Maurice), *drogues et produits chimiques.*

Jandin (Otto), *café-restaurant,* et boul. Beaumarchais, 33 et 35.

Mahon (H.) freres, *maladies du cuir chevelu,* et boul. Malesherbes, 9.

Mahon (de Molènes), *médecin.*

Molènes (de) pere, *médecin.*

Vilcoq (Emile), *lingerie.*

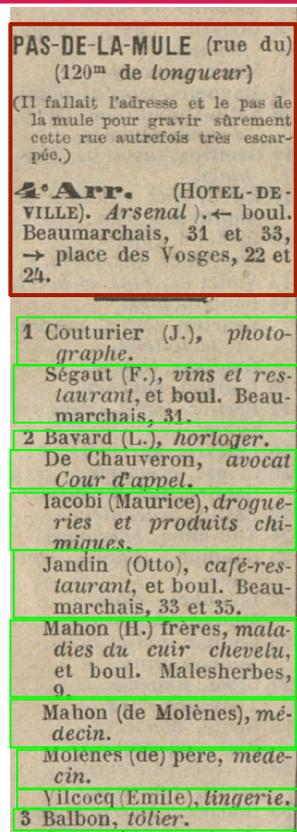
3 Balbon, *tôlier.*

Use cases: Legacy Address Directories

- Khemakhem et al. 2018c

Model	Annotated Data	F1-Score	
		Micro Average	Macro Average
Dictionary Segmentation	10 Pages	99.61	72.12
	7 training, 3 evaluation		
Dictionary Body Segmentation	319 Entries	98.61	95.7
	270 training, 49 evaluation		
Lexical Entry	208 Entries	90.31	91.36
	160 training, 48 evaluation		

Evaluation of the first three segmentation models

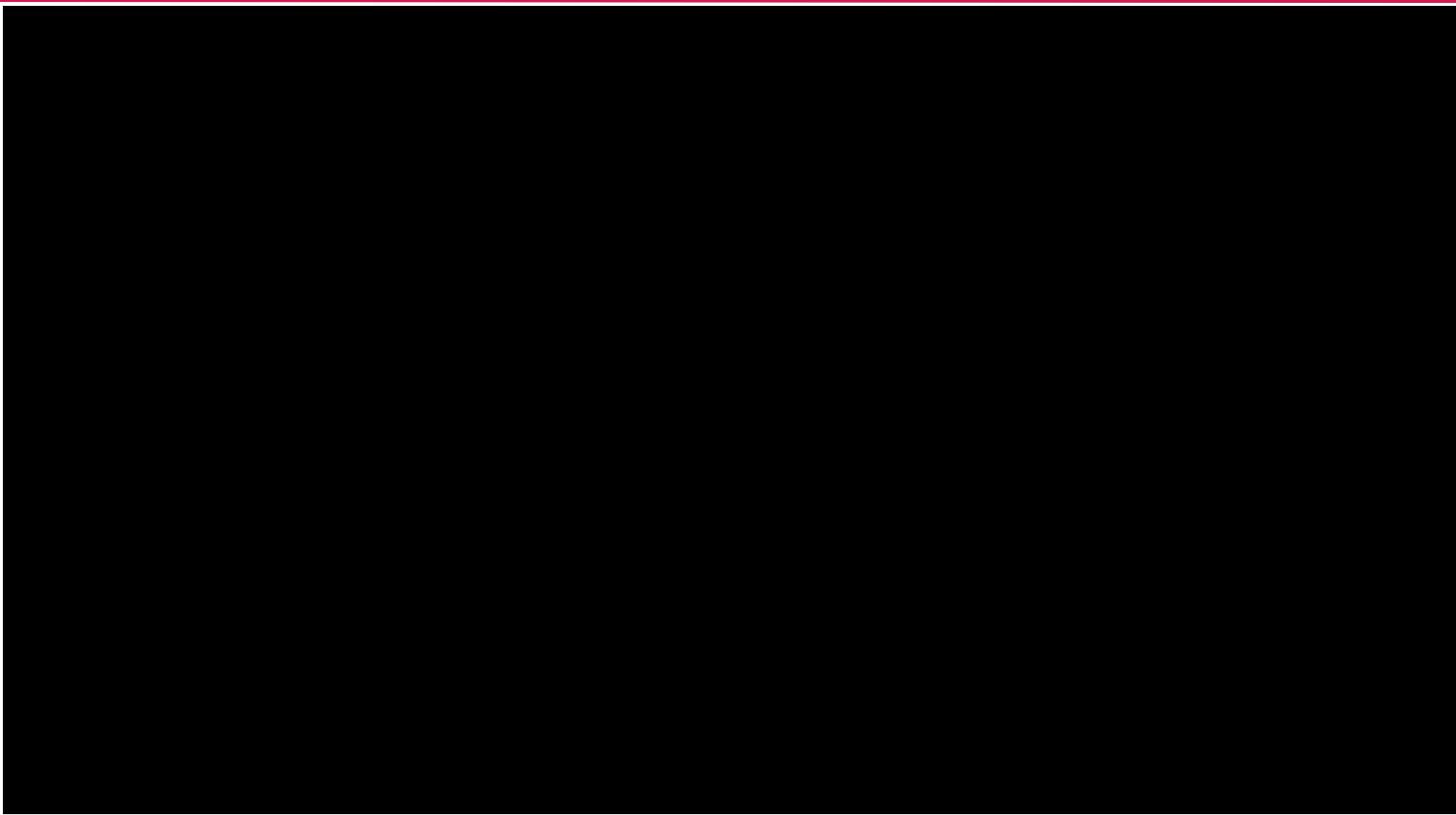


Source:

<https://gallica.bnf.fr/ark:/12148/cb32695639f/date19010101>

Use cases: Time Machine (Maps of Space and Time)

Source:
Frederic Kaplan's TEDx
Talk (June 2013)



Use cases: Wiegand Bibliography

- Lindemann and Khemakhem 2018

Titelstrecke V

21638 V.: Wozu bedarf ein Deutscher eines deutschen Wörterbuchs? In: Illustrierte Zeitung Nr. 1220 v. 17. 11. 1866, 327. [Presstext].

21639 v. B.: Duden – Das große Wörterbuch der deutschen Sprache in 6 Bänden. Über 500 000 Stichwörter und Definitionen: von A bis *Zyotoxizität*. Bibliographisches Institut, Mannheim. Dudenverlag. Preis je Band: 58 DM. In: Feld und Wald (Essen) v. 28. 1. 1983. [Presstext].

21640 *Vacatopoulou*, Anna: The 'Bank of English' and the 'British National Corpus': Evaluation from the Lexicographer's Perspective. Diss. [masch.] for M. A. in Lexicography. Univ. of Exeter. Exeter 1996.

21641 *Vacek*, Jaroslav: Russian Dictionaries of Indian Languages – a Note. In: Archiv Orientální 49. 1981, 66.

21642 *Vacek*, Jaroslav: A Computer Aided Dictionary of Tamil. In: Archiv Orientální 61. 1993, 317–319. [Rezensionsaufsatz].

Dazu Abstract in: LLBA 29. 1995, 2449.

21643 *Vachková*, Marie: Bilinguale Lexikographie und Wortbildung. Das adjektivische Suffix *-haft* in kontrastiver Sicht. In: Germanistica Pragensia 14. 1997, 143–150.

21644 *Vachková*, Marie: Das große deutsch-tschechische Wörterbuchprojekt im Kontext der zeitgenössischen deutsch-tschechischen Übersetzungsllexikographie. In: Kunzmann-Müller/Zielinski (Hrsg.) [...] 2002], 112–118.

21645 *Vachková*, Marie: Wortbildung und zweisprachiges Wörterbuch. In: Das Wort in Text und Wörterbuch [...] 2002], 119–126.

21646 *Vachková*, Marie: Der deutsch-tschechische Sprachvergleich im deutsch-tschechischen Übersetzungswörterbuch. In: Brücken. Germanistisches Jahrbuch Tschechien-Slowakei 11. 2003, 251–263.

21647 *Vaciago*, Paolo: Old English Glosses to Latin Texts: A Bibliographical Handlist. In: Medioevo e Rinascimento. Annuario del Dipartimento di Studi sul Medioevo e il Rinascimento dell' Università di Firenze 7. NS. 4 1993, 1–67. [Bibliographie].

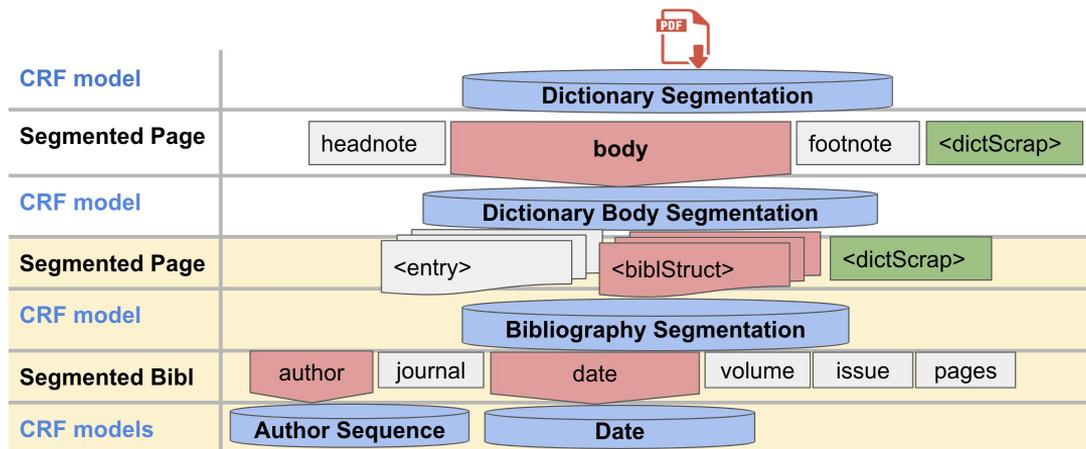
21648 *Vaciago*, Paolo: Towards a Corpus of Carolingian Biblical Glossaries. A Research in Progress Report. In: Les manuscrits des lexiques et glossaires [...] 1996], 127–144.

21649 *Vagianou*, Maria: Cartoon Humor in Children's Dictionaries and Reference Books and Its Effect on Learning and Vocabulary Retention. Diss. [masch.] for M. A. in Lexicography. Univ. of Exeter. Exeter 1996.

Bereitgestellt von | Universitätsbibliothek Hildesheim
Angemeldet
Heruntergeladen am | 04.05.18 14:03

Use cases: Wiegand Bibliography

- Lindemann and Khemakhem 2018



Titelstrecke V

21638 V.: Wozu bedarf ein Deutscher eines deutschen Worterbuchs? In: Illustrierte Zeitung Nr. 1220 v. 17. 11. 1866, 327. [Presstext].

21639 v. B.: Duden – Das groe Worterbuch der deutschen Sprache in 6 Banden. ber 500 000 Stichwortr und Definitionen: von A bis Zytotoxizitat. Bibliographisches Institut, Mannheim. Dudenverlag. Preis je Band: 58 DM. In: Feld und Wald (Essen) v. 28. 1. 1983. [Presstext].

21640 Vacatopoulou, Anna: The 'Bank of English' and the 'British National Corpus': Evaluation form the Lexicographer's Perspective. Diss. [masch.] for M. A. in Lexicography. Univ. of Exeter. Exeter 1996.

21641 Vacek, Jaroslav: Russian Dictionaries of Indian Languages – a Note. In: Archiv Orientalni 49. 1981, 66.

21642 Vacek, Jaroslav: A Computer Aided Dictionary of Tamil. In: Archiv Orientalni 61. 1993, 317–319. [Rezensionsaufsatz].

Dazu Abstract in: LLBA 29. 1995, 2449.

21643 Vachkova, Marie: Bilinguale Lexikographie und Wortbildung. Das adjektivische Suffix *-haft* in kontrastiver Sicht. In: Germanistica Pragensia 14. 1997, 143–150.

21644 Vachkova, Marie: Das groe deutsch-tschechische Worterbuchprojekt im Kontext der zeitgenossischen deutsch-tschechischen bersetzungslexikographie. In: Kunzmann-Muller/Zielinski (Hrsg.) [...] 2002], 112–118.

21645 Vachkova, Marie: Wortbildung und zweisprachiges Worterbuch. In: Das Wort in Text und Worterbuch [...] 2002], 119–126.

21646 Vachkova, Marie: Der deutsch-tschechische Sprachvergleich im deutsch-tschechischen bersetzungsworterbuch. In: Brucken. Germanistisches Jahrbuch Tschechien-Slowakei 11. 2003, 251–263.

21647 Vaciago, Paolo: Old English Glosses to Latin Texts: A Bibliographical Handlist. In: Medioevo e Rinascimento. Annuario del Dipartimento di Studi sul Medioevo e il Rinascimento dell' Universita di Firenze 7. NS. 4 1993, 1–67. [Bibliographie].

21648 Vaciago, Paolo: Towards a Corpus of Carolingian Biblical Glossaries. A Research in Progress Report. In: Les manuscrits des lexiques et glossaires [...] 1996], 127–144.

21649 Vagianou, Maria: Cartoon Humor in Children's Dictionaries and Reference Books and Its Effect on Learning and Vocabulary Retention. Diss. [masch.] for M. A. in Lexicography. Univ. of Exeter. Exeter 1996.

Bereitgestellt von | Universitatsbibliothek Hildesheim
 Angemeldet
 Heruntergeladen am | 04.05.18 14:03

Conclusion & Outlook

- New approach for structuring Lexical, Encyclopedic and Bibliographic information in entry based documents
- Standardised output to enable exchangeability and enrichment

- More feature engineering is required
 - Collect more varied data through GROBID-Dictionaries Workshop Series
- New techniques could be also tested for a hybrid processing

References

- **Sean Crist (2011)**. Processing the text of bilingual print dictionaries
- **Petra Bago. and Nikola Ljubešić (2015)**. Using machine learning for language and structure annotation in an 18th century dictionary. In Electronic lexicography in the 21st century: linking lexical data in the digital age
- **Mohamed Khemakhem, Luca Foppiano, Laurent Romary (2017)**. Automatic Extraction of TEI Structures in Digitized Lexical Resources using Conditional Random Fields. electronic lexicography, eLex 2017, Sep 2017, Leiden, Netherlands
- **Mohamed Khemakhem, Axel Herold, Laurent Romary (2018a)**. Enhancing Usability for Automatically Structuring Digitised Dictionaries. GLOBALEX workshop at LREC 2018, May 2018, Miyazaki, Japan. 2018
- **Mohamed Khemakhem, Laurent Romary, Simon Gabay, Hervé Bohbot, Francesca Frontini, et al.(2018b)**. Automatically Encoding Encyclopedic-like Resources in TEI. The annual TEI Conference and Members Meeting, Sep 2018, Tokyo, Japan
- **Mohamed Khemakhem, Carmen Brando, Laurent Romary, Frédérique Mélanie-Becquet, Jean-Luc Pinol (2018c)**. Fueling Time Machine: Information Extraction from Retro-Digitised Address Directories. JADH2018 "Leveraging Open Data", Sep 2018, Tokyo, Japan
- **David Lindemann and Mohamed Khemakhem (2018)**. Retro-digitizing and Automatically Structuring a Large Bibliography Collection. European Association for Digital Humanities (EADH) Conference, December 2018, Galway, Ireland

Thank you for your attention :)

Questions?