

**THE DIACHRONIC DEVELOPMENT OF COMBINING
FORMS IN SCIENTIFIC WRITING**

Katrin Menzel, Stefania Degaetano-Ortlieb*

*Corresponding author**


Bibliographic description:

Menzel, K. & Degaetano-Ortlieb S. (2017). The diachronic development of combining forms in scientific writing. In *Lege artis. Language yesterday, today, tomorrow. The journal of University of SS Cyril and Methodius in Trnava*. Warsaw: De Gruyter Open, 2017, vol. II (2), December 2017, p. 185-249. DOI: [10.1515/lart-2017-0016](https://doi.org/10.1515/lart-2017-0016) ISSN 2453-8035

Full version of the text can be downloaded from our official web site:

<http://www.degruyter.com/view/j/lart> or using DOI.

<p>Contact data Dr. Katrin Menzel Postdoctoral researcher, Department of Language Science and Technology at Saarland University, Campus A2.2, 66123 Saarbrücken, Germany, e-mail: k.menzel@mx.uni-saarland.de</p>		<p>Fields of interest</p> <p>Corpus linguistics, morphology, discourse analysis, translation studies, historical linguistics</p>
---	---	---

<p style="text-align: center;">Contact data</p> <p>Dr. Stefania Degaetano-Ortlieb</p> <p>Postdoctoral researcher, Department of Language Science and Technology at Saarland University, Campus A2.2, 66123 Saarbrücken, Germany e-mail: s.degaetano@mx.uni-saarland.de</p>		<p style="text-align: center;">Fields of interest</p> <p>Computational and corpus-based linguistics, data mining, applied linguistics and translatology, English scientific language, information density and linguistic encoding, evaluation/appraisal/stance.</p>
---	--	--

Résumé in English

Our study addresses the diachronic development of combining forms in English scientific texts over approximately 350 years, from the early stages of the first scholarly journals that were published in English to contemporary English scientific publications. Combining forms as bound lexical morphemes (e.g. *lith(o)- / -lith, graph(o)- / -graph, bio-, -lysis*) share some similarities with affixes and, at the same time, with base lexemes and parts of regular compounds. They seem to have always played a particularly important and productive role among English lexeme-formation elements in languages for special purposes, especially for the creation of new nouns. In this paper, we present a critical discussion of the category of combining forms as well as a case study that examines the role of selected combining forms in English scientific discourse. We use two diachronic corpora that consist of scientific texts from various disciplines from the middle of the 17th century onwards to the beginning of the 21st century – the Royal Society Corpus (RSC) and the Scientific Text Corpus (SciTex). Combining several lexical morphemes within single lexical items is a word

formation strategy that is particularly important for informational texts from scientific domains. What we primarily consider is the surprisal value of each unit. Surprisal is an information-theoretic notion related to the predictability and information density of text elements and measures the probability of a unit to occur in a given textual context. We present the insights that can be gained from considering surprisal values of combining forms and the elements, with which they co-occur in complex lexemes, measuring the probability of morphemes occurring together in specific time periods. The results of our case study have shown that a more predictive and conventionalized use of particular forms allows a more productive use of those forms in closely related or analogous grammatical contexts. Combining forms that are used as components of nouns in a rather predictable way, for instance, become easily productive in other word classes as well.

Key words: combining forms, morphology, history of scientific English, language for specific purposes, information density, corpus linguistics.

Résumé in German

In unserem Beitrag untersuchen wir die Entwicklung von Konfixen (*combining forms*) in englischen Wissenschaftstexten über einen Zeitraum von circa 350 Jahren, und zwar seit dem Aufkommen der ersten wissenschaftlichen englischen Zeitschriften bis hin zu zeitgenössischen Artikeln aus englischsprachigen wissenschaftlichen Fachzeitschriften. Die von uns betrachteten Konfixe sind gebundene lexikalische Morpheme (z.B. *lith(o)- / -lith*, *graph(o)- / -graph*, *bio-*, *-lysis*), die einige Gemeinsamkeiten sowohl mit Affixen als auch mit Basislexemen und Bestandteilen von regulär gebildeten Komposita haben. Sie spielen im Englischen besonders bei der Neubildung von Nomen in der Wissenschaftssprache schon seit langem eine besondere Rolle. In diesem Beitrag

präsentieren wir eine kritische Diskussion dieser Morphemkategorie in Abgrenzung zu anderen Wortbildungselementen. In einer Fallstudie beschäftigen wir uns detailliert mit ausgewählten Konfixen im englischen Wissenschaftsdiskurs. Es werden zwei diachrone Korpora bestehend aus wissenschaftlichen Texten verschiedener Disziplinen über den Zeitraum von Mitte des 17. Jhd. bis Anfang des 21. Jhd. für die Analyse herangezogen – das Royal Society Corpus (RSC) und das Scientific Text Corpus (SciTex). Die Verbindung mehrerer solcher Morpheme mit lexikalischer Bedeutungsfunktion innerhalb von Einzelwörtern ist eine wichtige Wortbildungsstrategie in wissenschaftlichen Fachtexten. In der Analyse legen wir einen Schwerpunkt auf den Surprisal-Wert ('Überraschungswert') der jeweiligen Einheiten. Surprisal als Begriff aus der Informationstheorie sagt etwas aus über die Vorhersagbarkeit und den Informationsgehalt von textuellen Bestandteilen und misst die Auftretenswahrscheinlichkeit dieser im jeweiligen textuellen Kontext. Es wird aufgezeigt, welche Erkenntnisse anhand dieser Werte für Konfixe und andere Morpheme, mit denen sie in komplexen Lexemen gemeinsam auftreten, gewonnen werden können. Hierbei wird bestimmt, mit welcher Wahrscheinlichkeit die untersuchten Morpheme in bestimmten Zeitperioden gemeinsam innerhalb von Lexemen auftreten. Die Ergebnisse zeigen, dass bei einem vorhersagbareren und konventionalisierteren Gebrauch bestimmter Formen diese produktiver in ähnlichen, bzw. analogen grammatischen Kontexten verwendet werden. Konfixe, die als Bestandteile von Nomen vorhersagbarer werden, finden beispielsweise dann auch produktiver als Elemente von Wörtern anderer Wortarten Verwendung.

Stichwörter: Konfixe (combining forms), Morphologie, Geschichte der englischen Wissenschaftssprache, Fachsprache, Informationsdichte, Korpuslinguistik.

Résumé in French (translated by Olivier Landeville)

Notre article a pour objet d'étudier l'évolution des confixes (*combining forms*) dans des textes scientifiques rédigés en anglais sur une période s'étalant sur 350 ans environ, depuis la publication des premiers journaux scientifiques anglophones jusqu'aux articles contemporains extraits de revues scientifiques publiées en anglais. Les confixes sont des morphèmes lexicaux liés (p. ex. en anglais *lith(o)- / -lith, graph(o)- / -graph, bio-, -lysis*) qui ont certaines similarités avec les affixes et, dans le même temps, avec les lexèmes de base et les éléments d'autres mots composés. Ils semblent jouer déjà depuis longtemps dans la langue anglaise un rôle particulier, notamment lors de la formation de nouveaux substantifs dans le langage scientifique. Dans cet article, nous offrons un débat critique sur les confixes en tant que catégorie de morphèmes. Dans une étude de cas, nous examinons le rôle de certains confixes dans le discours scientifique en anglais. Dans le cadre de cette analyse, deux corpus diachroniques composés de textes scientifiques issus de différentes disciplines et publiés du milieu du 17^e siècle au début du 21^e siècle sont utilisés: le Royal Society Corpus (RSC) et le Scientific Text Corpus (SciTex). La combinaison de plusieurs morphèmes lexicaux de ce type permettant de former des mots individuels complexes constitue une stratégie de création lexicale importante dans les textes scientifiques. Lors de l'analyse, l'accent est mis sur les valeurs de prédictibilité (*surprisal*), c'est-à-dire sur la probabilité qu'un élément apparaisse dans un contexte textuel donné. Issu de la théorie de l'information, ce concept porte sur la prédictibilité et la densité d'information d'éléments textuels. Nous présentons les conclusions qui peuvent être tirées en prenant en compte ces valeurs pour les confixes et autres morphèmes avec lesquels ils apparaissent conjointement dans des lexèmes complexes, tout en étudiant la probabilité avec laquelle les morphèmes analysés apparaissent dans des lexèmes sur des périodes spécifiques. Les résultats de notre étude de cas montrent que si l'usage de certaines formes est plus prévisible et plus

conventionnalisé, celles-ci seront utilisées de manière plus productive dans des contextes grammaticalement similaires et analogues. Les confixes utilisés par exemple en tant que composants de noms de manière plutôt prévisible deviennent facilement productifs dans d'autres catégories de mots.

Mots-clés: confixes (combining forms), morphologie, histoire de l'anglais scientifique, langue de spécialité, densité d'information, linguistique de corpus.

Résumé in Russian (translated by Ekaterina Lapshinova-Koltunski)

В данной статье рассматривается историческое развитие компонентов сложных слов в английских научных текстах на протяжении приблизительно 350 лет, начиная от ранних научных публикаций на английском языке заканчивая современным английским. Компоненты сложных слов всегда являлись важными и продуктивными процессами формирования лексических единиц в специальных языках, особенно в формировании новых существительных. В данной статье представлен критический анализ категории компонентов сложных слов на примере избранных компонентов в английском научном дискурсе. Мы используем два диахронных корпуса научных текстов из разных дисциплин, созданных на протяжении многих лет, начиная с 17-го и заканчивая 21-м веком – Royal Society Corpus (RSC, Корпус Королевского Научного Общества) и Scientific Text Corpus (SciTex, корпус научных текстов на английском языке, созданных в Дармштадте и Саарбрюкене). Для информативных текстов научного дискурса характерным способом словообразования является комбинирование лексических морфем в одну лексическую единицу. В первую очередь, нас интересует surprisal (не/предсказуемость) каждой единицы, то есть вероятность появления данной единицы в данном контексте. Мы представляем результаты исследования

величины surprisal компонентов сложных слов и элементов, с которыми они сочетаются в сложных лексемах, измеряя вероятность сочетаемости морфем в определённые периоды времени. Результаты данного исследования показывают, что более предсказуемое и стилизованное использование определённых форм позволяет более продуктивное использование этих форм в тесно связанных контекстах.

Ключевые слова: компоненты сложных слов, морфология, история английского языка, Язык для специальных целей (LSP), Информационная плотность, корпусная лингвистика.

Article was received by the editorial board 27.06.17.

Reviewed 23.07.17. and 21.09.17.

Similarity Index 3%