

TDR

SYSTÈME DE RÉOLUTION DE L'ANAPHORE PRONOMINALE ADVERBIALE (PAAR)

Hermann SOCKENG
Ingénieur en Calcul Scientifique
hermannsockeng@gmail.com

Nelson Leroi SOFAC
Doctorant en Linguistique Informatique
sofacleroi@gmail.com

November 18, 2023

1 Description

L'objectif de ce projet est de lier deux parties d'un discours. L'une des parties est un mot (adverbe pronominal, abrégé en PA) et l'autre partie est une phrase précédant le mot (l'antécédent).

Le PA peut être relié aussi bien à la phrase précédente (antécédent) qu'à la phrase qui le suit (postcédant). Lorsque le PA est relié à la phrase précédente, on l'appelle **anaphore**, et lorsqu'il est relié à la phrase suivante, on l'appelle **cataphore**.

2 Objectif

L'objectif de ce projet est d'identifier automatiquement les PA dans un paragraphe donné et de déterminer s'ils font référence à un **antécédent** ou à un **postcédant**.

3 Sources d'informations existantes

Le processus consistant à déterminer si un PA est un antécédent ou un postcédant a été modélisé par un linguiste au moyen d'un algorithme. Les étapes comprennent :

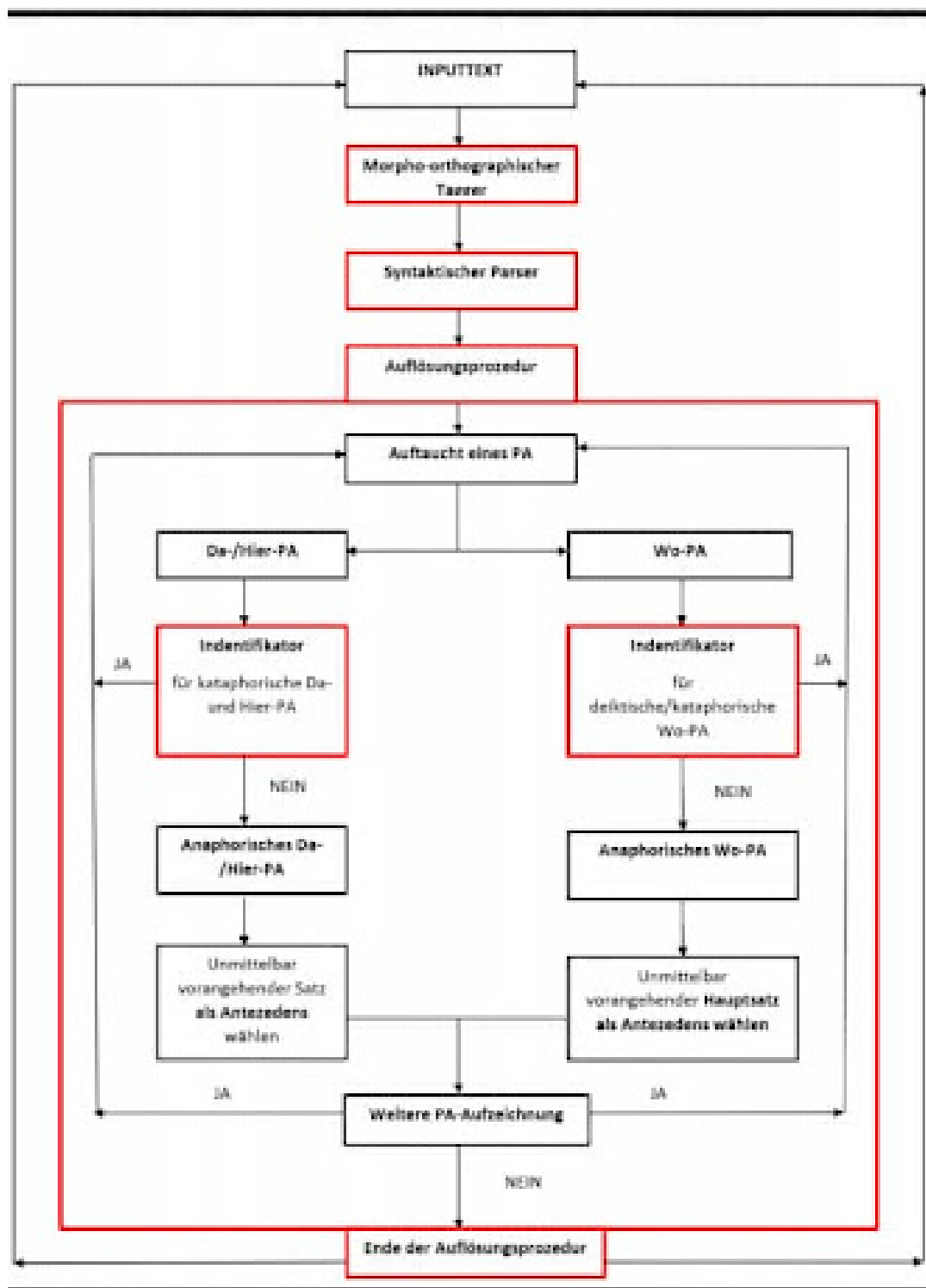
1. Algorithme de résolution de l'anaphore pronominale adverbiale.

Cet algorithme établit le lien définitif entre le PA et son antécédent. Il requiert l'utilisation des algorithmes suivants :

1.1. Un algorithme qui détecte si le PA a un postcédant, c'est-à-dire s'il s'agit d'une cataphore. Si c'est le cas, on passe au PA suivant ; sinon, on considère ce PA comme ayant un antécédent (l'anaphore). On distingue deux formes de cet algorithme : l'une pour les **Da-/Hier-PA**, qui identifie les PA suivis d'au plus 3 mots, d'une virgule et d'un des mots suivants : **dass**, **ob**, et les **W-Satz** (environ 40). L'autre forme identifie les PA suivis d'au plus 1 mot et d'une virgule ou de deux points.

1.2. Wo-PA, qui identifie les PA en début de phrases.

L'algorithme général est illustré dans la figure 1.



**Abbildung 7: Endgültiger Algorithmus zur Auflösung von
Pronominaladverbialanaphern**

4 Aspects méthodologiques

La première étape consiste à **tokenizer** le texte en phrases et en mots afin d'identifier les PA. Cela nécessite l'utilisation d'un logiciel approprié, notamment **Python**, qui est l'un des langages de programmation les plus utilisés en linguistique computationnelle et en traitement du langage naturel (NLP).

Compréhension du problème Le problème consiste à classifier automatiquement les ad-
verbes pronominaux d'un texte en langue allemande, comme décrit dans les sections précédentes.

Approche analytique Nous nous basons sur l'algorithme de résolution de l'anaphore pronom-
inal adverbiale fourni par le linguiste, ainsi que sur les méthodes prédéfinies en Python, notam-
ment le framework NLTK (Natural Language Toolkit).

Données nécessaires et collectées Au départ, nous avons seulement besoin de la liste
des adverbes pronominaux. Une fois les méthodes implémentées, les données de test seront
fournies par le linguiste.

Évaluation de l'algorithme Les résultats obtenus sur les données de test fournies par le
linguiste seront consignés dans des graphiques.

Déploiement du modèle et retour d'expérience À venir !

5 Participation et responsabilité des parties

Le linguiste et l'ingénieur en calcul scientifique travaillent en collaboration. Le linguiste
est principalement chargé de décrire et de présenter les algorithmes basés sur l'exploitation
des principes linguistiques, tels que l'algorithme de filtrage des cataphores et l'algorithme de
résolution globale. L'ingénieur en calcul scientifique, quant à lui, s'approprie ces méthodes et
les implémente en langage informatique.