Emerging NPIs: The acquisition of Dutch hoeven 'need'

Lin, J.; Weerman, F.; Zeijlstra, H.

Link to publication

Jing Lin*, Fred Weerman and Hedde Zeijlstra

Emerging NPIs: The acquisition of Dutch *hoeven* 'need'

Abstract: Dutch modal verb *hoeven* 'need' is a Negative Polarity Item (NPI) because of its restricted distribution to certain negative contexts only. By investigating the distribution of this NPI in child Dutch, the paper explores a solution to a learnability problem raised by the existence of NPIs: how can a child acquire the limited distribution of an NPI in the absence of both direct and indirect negative evidence? Corpus data collected through CHILDES confirm children's employment of a conservative widening learning strategy to solve the learnability problem. This strategy entails that children start out with the strictest assumption of *hoeven*, exhibiting a lexical dependency with the negative marker *niet* 'not', and weaken the assumption down to a less rigid reanalysis of this NPI, associated with an abstract negation in its underlying syntactic representation. The initial learning process turns out to be distribution-based only, i.e., without presuming any innate knowledge of NPIs and their restricted occurrences. However, distributional properties alone are not sufficient for children to reanalyze the NPI. Children's linguistic knowledge of negative indefinites as exhibiting a decomposable negation plays a crucial role in the subsequent reanalyzing process. The reanalysis emerging shortly after age four signifies exactly how adult speakers analyze the NPI, also explaining *hoeven*'s strength as a polarity item.

Keywords: acquisition, corpus study, Dutch *hoeven*, learnability, Negative Polarity Items (NPIs)

1 Introduction

Negative Polarity Items (NPIs) are lexical items that can only survive in certain kinds of negative contexts (see Ladusaw 1979; McCawley 1988; etc.). Dutch modal

---

*Corresponding author: Jing Lin: University of Amsterdam. E-mail: J.Lin@uva.nl
Fred Weerman: University of Amsterdam.
Hedde Zeijlstra: University of Göttingen.

verb *hoeven* 'need' is an NPI (Zwarts 1981, 1986, 1998; Van der Wouden 1997; Hoeksema 1997, 2000). This is shown in examples below, in which *hoeven* is only grammatical under the scope of the sentential negative marker *niet* 'not' in (1), or negative indefinite *geen* 'no(ne)' or *niets* 'nothing' in (2), or semi-negative expressions such as *nauwelijks* 'seldom' in (3a) and exclusive adverbs like *alleen* 'only' in (3b).[1]

(1) *Wij <u>hoeven</u> vandaag *(**niet**) te werken.*
    we   need   today   not   to   work
    'We do not have to work today.'

(2) a. **Geen**/**Een arts <u>hoeft</u> vandaag te werken.*
      no/a   doctor   needs   today   to   work.
      'No doctor has to work today.'
   b. *Jan <u>hoeft</u> **niets**/*iets voor zijn dochter te kopen.*
      John   needs   nothing/something   for   his   daughter   to   buy
      Lit.: 'John has to buy nothing for his daughter.'
      'John does not have to buy anything for his daughter.'

(3) a. *Jan <u>hoeft</u> *(**nauwelijks**) te koken.*
      John   need   seldom   to   cook
      'John seldom has to cook.'
   b. *Jan <u>hoeft</u> *(**alleen**) in het weekend te koken.*
      John   need   only   on   the   weekend   to   cook.
      'John only has to cook on the weekends.'

NPIs form a cross-linguistic phenomenon (Horn 1989; Haspelmath 1997), and may vary cross- and intra-linguistically in strength depending on the types of negative contexts that may license them (Zwarts 1986, 1995; Nam 1994; Van der Wouden 1997; Hoeksema 2000). Since the ungrammaticality of an NPI in non-licensing contexts is not necessarily inferred by the absence of such constructions in the input, the existence of NPIs raises an important problem for language learners. How can children detect an NPI's restricted distribution in the absence of substantial or reliable negative feedback? The fact that NPIs vary in their strengths makes the learnability problem even more complex: how and why are language learners able to acquire, in the absence of negative evidence, exactly which types of contexts can and cannot license an NPI? Although from the per-

---

1 For a systematical representation of *hoeven*'s distribution in Dutch as an NPI, the reader is referred to Section 6.

spective of learnability (Gold 1967; Chomsky 1972; 1975, 1981; Goldsmith 1976, 1980; Pinker 1979; among others) the existence of NPIs poses hard problems, surprisingly little research has been done on this topic. To the best of our knowledge, the only existing work on the topic is limited to Dutch NPIs *hoeven* 'need' and *meer* 'more' in Van der Wal (1996), English NPI *any* in Tieu (2009, 2011) and Mandarin NPI *shenme* 'a (thing)' in Lin (2011a), Lin and Zeijlstra (2012).[2]

In order to solve the learnability problem, the current study explores the acquisition of the NPI *hoeven* by analyzing spontaneous speech data of 53 monolingual Dutch children. The paper shows that the learnability problem can only be circumvented once a *conservative widening learning strategy* is adopted. We provide evidence for the following acquisitional process. First, children start out with a strict constructional analysis of the target NPI (in this case: *hoeven*). This is based on positive evidence that *hoeven* has a lexical dependency with the negative marker *niet*. In the subsequent stage, children extend this initial assumption according to language input and reanalyze *hoeven* as lexically associated with an abstract semantic negation NEG in its underlying syntactic representation. Moreover, we show that this reanalysis turns out to be the adult analysis of the target NPI and consequently explains why *hoeven* is a relatively strong NPI.

The paper first outlines the learnability problem in Section 2. Then, in Section 3, it illustrates the conservative widening learning hypothesis and motivates three predictions for the development of child Dutch. Section 4 introduces the methodology and Section 5 presents our research results. In Section 6, we provide an explanation for the attested acquisitional pathway of the NPI. Section 7 compares our approach with two alternative explanations: Van der Wal's approach (1996) and a lexical approach following Mintz et al. (1995, 2002) and Mintz (2002, 2003). Section 8 concludes the paper and presents suggestions for further research.

2 The learnability problem

NPIs such as *hoeven* exhibit a restricted distribution to certain negative contexts only (see (1) to (3)). This suggests that children, in order to achieve the target

---

2 Drenhaus et al. (2005), Drenhaus et al. (2006a, 2006b), Pablos et al. (2011), Saddy et al. (2004), Schutte (2006) and Xiang et al. (2009) investigate how the participants – both adults and children – process grammatically and ungrammatically used NPIs in Basque, English, German and Greeks. However, none of these studies focuses on the acquisition of NPIs by language acquiring children and the learnability problem remains unaddressed.

analysis of an NPI, must obtain the knowledge of not only exactly which types of negative contexts can license it but also in exactly which types of negative contexts it is not allowed to appear. Acquisition of possible licensing contexts for an NPI is easily supported by the presence of positive evidence. This is because positive evidence refers to input data containing grammatical constructions in the target language (Pinker 1995) that is reliable and available to all language learners in all developmental stages (Pinker 1984, 2013; Marcus 1993; among others). On the contrary, to acquire the knowledge of what are impossible licensing contexts for *hoeven* appears to be problematic: the absence of unlicensed NPIs in the input does not necessarily indicate any ungrammaticality. In such a situation, we would expect children to be systematically confronted with explicit information on the ungrammaticality of unlicensed NPIs, i.e., direct negative evidence (Baker and McCarthy 1981; Marcus 1993; among others). However, as is widely argued, children acquire language solely from positive evidence (Braine 1971; Chomsky and Lasnik 1977; Baker and McCarthy 1981; Grimshaw 1981; Pinker 1984, 1995, 2013; Berwick 1985; Manzini and Wexler 1987; Gropen et Mal. 1991; Marcus et al. 1992; Marcus 1993). Negative evidence, such as corrective feedback from parents, or an explicit denial of a child's ungrammatical utterance, is either absent (Brown and Hanlon 1970; Braine 1971; Pinker 1984, 1995, 2013; Gropen et al. 1991; Marcus et al. 1992; Marcus 1993; Cowie 1997; Boyd and Goldberg 2011; Goldberg 2011; among others),[3] or "(mercifully) uncommon" (Boyd and Goldberg 2011: 56). Marcus (1993) also claims that negative evidence is not reliable as it is not of sufficient quantity and quality; neither is it available for all language learners in all acquisitional stages for all types of ungrammatical errors.[4] Moreover, language learners do not seem to benefit from corrective feedback, even if there is any (McNeill 1966; Stromswold 1994; MacWhinney 2002; Boyd and Goldberg 2011; among others). In the absence of substantial and influential negative evidence, overgeneralization errors appear to be unavoidable (Gold 1967; Braine 1971; Pinker 1979, 1995, 2013). Children are therefore ex-

---

3 See also MacWhinney (2002) for a review of the literature in this respect.

4 See a related discussion in Marcus (1993) on the affectedness of different types of corrective feedback in the input on children's acquisition of the target language. Actually, the only possible form of corrective feedback in a process of language acquisition is what Marcus (1993) terms "the noisy feedback", i.e. "certain discourse patterns that differ in frequency depending on the grammaticality of children's utterances" (Marcus 1993: 53). However, as the author argues, such feedback is too unreliable and unsystematic to have any influence on children's acquisition of the knowledge of what is ungrammatical in their target grammar (see also Boyed and Goldberg 2011, among others).

pected to overuse NPIs in non-licensing environments. But if overgeneralization errors indeed occur, how are children supposed to unlearn these errors in the absence of influential systematic corrective feedback or explicit information on unlicensed NPIs?

One possible mechanism that may help language learners to recover from overgeneralizations is that of *statistical pre-emption*. This learning mechanism does not require language learners to be able to benefit from direct negative evidence, since it only relies on indirect negative evidence (Bates and MacWhinney 1987; Clark 1987; Goldberg 1993, 1995, 2006, 2009, 2011; Marcotte 2005; Foraker and McElree 2007; Boyd and Goldberg 2011). Indirect negative evidence (cf. Chomsky 1981) refers to information about which constructions or forms are absent from the language input, and is a third type of evidence in language acquisition, in addition to positive evidence and direct negative evidence. In order to benefit from indirect negative evidence, language learners must make the following inference: X is not part of the target language if X is absent from the input. For instance, every time English children (would like to) use *asleep cats, they hear a semantically and pragmatically related alternative construction *sleeping cats* in the input. By their employment of indirect negative evidence, children can eventually establish a rule that the adjective *asleep* cannot be attributively used according to the target grammar. Contrary to direct negative evidence, indirect negative evidence is reliable and available to language learners regardless of their age or the type of grammatical phenomena (cf. Marcus 1993). This suggests that the absence of *asleep cats in the input can function as a cue for the ungrammaticality of attributively used *asleep* in adult English. The presence of *sleeping cats* preempts therefore the overused *asleep cats. The mechanism of statistical pre-emption requires two conditions. The first is the obligatory existence of the alternative form in the input, i.e., *sleeping cats* in this example; and the second is that the frequency of this alternative is much more than zero whereas that of the overused form, i.e., *asleep cats, is always zero.

The statistical mechanism could explain how children can unlearn overgeneralization errors at both the syntactic and morpho-syntactic level (see Boyed and Goldberg 2011 for an overview of related literature). Nevertheless, this mechanism cannot account for how children could ever recover from their overgeneralization errors of NPIs, since the first condition required by statistical pre-emption would never be satisfied in such cases. For instance, Dutch does not seem to exhibit a semantically or pragmatically alternative form of *hoeven* in a complementary distribution. At first sight, the Positive Polarity Item (hereafter PPI) *moeten* 'must' (cf. Iatridou and Zeijlstra 2010) may appear to be such an alternative modal, pre-empting the NPI *hoeven*; but there are contexts in Dutch that may even license both the NPI and the PPI: under the

scope of an exclusive adverb in (4) and a quasi-negative expression in (5).[5]

(4) a. ***Alleen*** *Jan* <u>*hoeft*</u> *te rijden.*
      only John needs to ride
      'Only John has to drive.'
   b. ***Alleen*** *Jan* <u>*moet*</u> *rijden.*
      only John must ride
      'Only John has to drive.'

(5) a. ***Niet iedereen*** <u>*hoeft*</u> *naar de WC te gaan.*
      not everybody needs to the toilet to go.
      'Not everybody has to go to the toilet.'
   b. ***Niet iedereen*** <u>*moet*</u> *naar de WC gaan.*
      not everybody must to the toilet go.
      'Not everybody has to go to the toilet.'

Examples in (4) and (5) clearly show that in the particular case of the acquisition of NPIs, indirect negative evidence is absent as well. However, as pointed out by one of the reviewers, language learners might still be able to appropriately restrict the distribution of two items by performing (a form of) statistical pre-emption,

---

5 The absence of any semantically or pragmatically related alternative forms of NPIs is not restricted to Dutch. English NPIs lack pre-empting forms as well because there are contexts in English that may allow both NPIs and non-NPIs to appear. This is illustrated in the examples below, in which NPIs are marked in bold.

(1) *Everyone who wants to have **any**/some/Ø coffee should go to the front desk.*

(2) *Has John finished **yet**/already?*

(3) *If you **ever**/sometimes feel sad, you can even call me at work.*

Similar observations are also made for Mandarin NPI *shenme* 'a (thing)', as both the NPI indefinite and its plain counterpart *yi ben* 'one-CL' can survive in the following contexts. Bare NPs are also allowed in these contexts.

(4) *Yuehan **shenme**/yi ben/Ø shu dou mei mai.*
     John *shenme*/one-CL/Ø book all not buy
     'John did not buy any book(s) at all.'

(5) *Yuehan zai kan **shenme**/yi ben/Ø shu ma?*
     John at read *shenme*/one-CL/Ø book Q-marker
     'Is John reading (a/some/any) book(s)?'

(6) *Yuehan haoxiang zai kan **shenme**/yi ben/Ø shu.*
     John probably at read *shenme*/one-CL/Ø book
     'John is probably reading (a/some) book(s).'

because they can sometimes tease apart environments in which the two items are truly complementary from those in which they are non-complementary. A case in point could be the English *some–any* dichotomy, which allows distributional overlap between the two in the same contexts where both *hoeven* and *moeten* can occur. Nevertheless, children cannot depend on an alternative item of a modal NPI to detract from their overuse of the NPI – whether or not this NPI has a truly complementary distribution with an alternative PPI – because Dutch NPI *hoeven* is not the only modal verb that necessarily takes scope under negation. As observed by Iatridou and Zeijlstra (2013), among others, all polarity-insensitive modal verbs take scope under negation as well. On the other hand, the PPI modal *moeten* always has a wide scope over negation (see, e.g., Iatridou and Zeijlstra 2013). If children would indeed rely on the existence of the PPI *moeten* to acquire the target modal *hoeven*, then it would suffice for them to assume that *hoeven* is a polarity-insensitive modal instead of an NPI modal. In fact, several languages exhibit a universal PPI modal next to a polarity-insensitive one. For instance, Greek *prepi* 'must' is a PPI, but its dual *chriazete* 'have to' is polarity neutral. This means that NPIs cannot be acquired solely by relying on their counterpart PPIs. Consequently, a statistical learning mechanism, such as *the (Constrained) Statistical Learning Hypothesis* (Saffran 2002, 2003; Romberg and Saffran 2010; Thiessen 2011) cannot provide a proper explanation for how language learners are supposed to unlearn overused NPIs in the absence of direct negative evidence.

The literature so far also provides other possible strategies that explain why in an actual process of language acquisition children can sometimes recover from disastrous overgeneralization errors – even in the absence of direct negative evidence. These strategies, being either learned or innate, concern "a general pragmatic mechanism or a linguistic specific mechanism" (Marcus 1993: 55) (see also Braine 1971). One example is a universal linking rule called *object affectedness* (Gropen et al. 1991) that links the argument specified as having a certain theta-role in a verb's semantic representation to the grammatical object. As discussed in Gropen et al. (1991), this universal linking rule explains how English children can unlearn their own erroneous constructions (see also Marcus 1993 and Pinker 2013). For example, in an utterance as **I filled water into the glass*, children overgeneralize the argument structure of *pour*, as in *I poured water into the glass*. However, they unlearn the overgeneralization once positive evidence leads them to acquire the correct semantic representation of the verb *fill*. Another example of internal mechanisms is a hypothesis of *inflectional blocking* (Marcus et al. 1992), also known as the *unique entry* principle in Pinker (1984). This hypothesis explains how children are able to unlearn overused regular morphemes to irregular word stems, just like overgeneralization errors of the regular past tense morpheme *-ed* to irregular verbs in English, resulting in erroneous forms such as

go-ed, or even *went-ed* and *make-d* (see also Marcus 1993; Clark 1995; Pinker 2013).

The two examples outlined above represent children's overgeneralization errors at the interface between (word) semantics and syntax (Gropen et al. 1991; Marcus 1993; Boyed and Goldberg 2011; Goldberg 2011; Pinker 2013; among others) and at the morpho-syntactic level (Bowerman 1982; Pinker 1984, 1995; Clark 1987, 1995; Marcus 1993; among others). Although both the universal linking rule and the inflectional blocking hypothesis are, naturally, too specific to apply to the particular case of overgeneralization of NPIs, we do not see how a more general mechanism in the same spirit could do so either.

So far, we have elaborated the learnability problem of NPIs and argued why language learners appear to suffer from overused NPIs that cannot be unlearned in the absence of both direct and indirect negative evidence. Since every typically developing child successfully acquires his/her first language, however, the learnability problem described here must be solvable. By focusing on the Dutch NPI *ho-even*, this paper explores a solution to the learnability problem of NPIs in general.

3 Hypothesis and predictions

The learnability problem of NPIs due to the absence of negative evidence can only be solved once a learning model is adopted that relies on positive evidence only. One possible learning strategy in this respect is *conservative widening* (after Van der Wal 1996; see also Manzini and Wexler 1987; Koster and Van der Wal 1996; Tieu 2009, 2011; Snyder 2010). Conservative widening is a general learning mechanism hypothesized for language acquisition that can be best defined in terms of the Subset Principle (Manzini and Wexler 1987): "Briefly, the subset principle demands that a learning procedure should guess the narrowest possible language, consistent with positive evidence seen so far. By hypothesizing as narrow a target language as possible, the acquisition procedure is protected from disastrous overgeneralization" (Berwick and Weinberg 1986: 233). Although a conservative widening learning mechanism can prevent children from making overgeneralization errors, the question is how children are supposed to make the first step in analyzing an NPI. Instead of assuming any innate linguistic knowledge, we take children's establishment of their initial analysis to be input-based only in a similar way as proposed for category learning via *a distributional approach* (Cartwright and Brent 1997; Mintz et al. 1995, 2002; Mintz 2002; Redington et al. 1998). We adopt Mintz (2002, 2003) and Mintz et al. (2002) in that a *distribution-based* learning mechanism plays a crucial role in early language acquisition. In the absence of innate knowledge about NPI-hood, children's first attempt to analyze a

target NPI can only be guided by investigating positive evidence available in the beginning of acquisition in terms of distributional properties.

The hypothesis discussed above predicts the acquisition of NPIs to exhibit several developmental stages. Children start with the strictest possible analysis of their target language, based on distributional properties of that limited input data available in the acquisitional onset. This analysis may not be identical to that of an adult speaker, but it is at least compatible with all the input data a child has perceived and analyzed so far. However, such a rigid analysis can be easily falsified by more input data processed in a succeeding stage.[6] This can eventually lead language learners to weaken down the initial analysis to construct a re-analysis explaining the input data perceived and processed in both stages. Such an iterative process continues until at a certain stage of language acquisition children may establish an analysis that explains all input data throughout the whole process of language development. This conservatively widening pathway in acquisition, as schematically illustrated below, can be best described by the Subset Principle: the set of the output of an analysis at a certain stage is always a subset of the set of the output generated by its reanalysis in the subsequent stage.

Table 1: An acquisitional process under the hypothesis of conservative widening

Stage 1:	limited input data available in the onset	→	initial analysis A1
Stage 2:	input data that falsify A1	→	reanalysis of A1: A2
Stage 3:	input data that falsify A2	→	reanalysis of A2: A3
…	…	…	…
Stage n:	Input data that falsify A(N-1)	→	reanalysis of A(N-1): AN = ATARGET

The hypothetical acquisitional process illustrated in Table 1 leads to three predictions for the distribution of *hoeven* in child Dutch development.

First, we expect children not to overgeneralize the NPI in non-licensing conditions in any developmental stage. The conservative widening strategy relies merely on positive evidence. Consequently, it is impossible for children to produce utterances of overused *hoeven* as such utterances are absent in the language input.

---

6 The question may now arise as to why children in an earlier stage ignore these falsifying data. We do not assume that the quality or quantity of the input changes over time, but rather that the amount of input evidence that language learners are able to process and analyze in an initial stage is more limited. This may be a consequence of children's small capacity of processing and analyzing input data, which improves over the course of language development; it may also be the case that it takes time for children to collect sufficient frequency data to falsify an initial analysis.

Second, we predict the distribution of *hoeven* to be more restricted in early than in late child Dutch. This is because the set of the output of *hoeven*'s analysis is a subset of the set of the output of its reanalysis in a subsequent stage, as illustrated in Table 1.

Third, since we do not assume any inherent awareness of *hoeven* being an NPI, contrary to Van der Wal (1996) as will be discussed in Section 7, we expect *hoeven*'s distributional pattern in the input to be important in children's first attempt to analyze the NPI. Following Mintz (2002, 2003) and Mintz et al. (2002), our last prediction is that *hoeven* in early child Dutch is restricted only to the type(s) of negative contexts that is/are most frequently attested as its licenser(s) in child-directed speech.

4 Methodology

In order to examine the three predictions from the perspective of a conservative widening learning strategy, we conducted a corpus study in the CHILDES database (MacWhinney 2009), in which spontaneous speech data of Dutch children in the following subcorpora were investigated: *BolKuiken* (Bol and Kuiken 1990),[7] *CLPF* (Fikkert 1994; Levelt 1994), *Groningen* (Wijnen and Bol 1993), *VanKampen* (Van Kampen 1994) and *Wijnen* (Wijnen 1988, 1992; Wijnen and Elbers 1993). Altogether, 633 CHAT files of 53 monolingual Dutch children between the ages of one and five were analyzed. Because only two of these children were longitudinally recorded throughout the period of investigation, i.e., Sarah and Laura in *VanKampen* (Van Kampen 1994), we opted for a cross-sectional analysis. To provide an overview of the distribution of *hoeven* in child Dutch development, three stages were divided based on the biological age of the child at the moment of recording: Stage I (age: 1–3), Stage II (age: 3–4) and Stage III (age: 4–5).[8]

---

7 Only typically developing children recorded in this subcorpus were included.
8 The current study chose the chronological age at recording as an indication of different developmental stages. Alternatively, as proposed by one of the reviewers, Dutch children's development of negation may also indicate different stages in the acquisition of NPIs. However, a potential difficulty of using the development of negation as an indicator is that children's production (of negation) does not always equal their knowledge (of negation). Another alternative would be to use the Mean Length of Utterance (MLU) in either morphemes or words to benchmark children's acquisitional stages. MLU in either morphemes or words is a useful tool for distinguishing atypically developing children from their typically developing counterparts (see, e.g., Bishop and Adams 1990; Nippold 1990; Nippold and Schwarz 2002) or to indicate language proficiency of children in relatively early language development, i.e., before approximately age four (Leonard 2000, among others). Since all our participants were typically developing children up to the age of five, we did not employ MLU in either morphemes or words in the current study.

The exact procedure of the investigation is as follows. The frequency of the target NPI in three inflected forms of the present tense and its infinitive form was first counted by employing the freq-command of the CLAN program in CLAN CHAT files, namely *hoef* (HOEVEN-1SG), *hoeft* (HOEVEN-2/3SG) and *hoeven* (HOEVEN-PL/INF). Subsequently, each utterance containing *hoeven* was analyzed for both its licensing status and licensing condition by using the kwal-command. Generally, three lines of context proceeding and following an utterance of *hoeven* were analyzed as well by adding "+w3" and "−w3" to the command. If a total of six lines of context were not sufficient to evaluate *hoeven*'s licensing status and licensing condition, more contextual data was checked manually.

5   Results and analysis

Children's utterances containing the target NPI were divided into three categories depending on licensing status, i.e., the category of *adult-like*, *non-adult-like* and *unclear situation*. The adult-like category contained utterances of *hoeven* that are also commonly produced by Dutch adult speakers. *Hoeven* in such utterances was either *linguistically licensed* by a proper linguistic licenser, such as the negative quantifier *geen* 'no(ne)' in (6); or *pragmatically licensed*, for instance, in a contrastive context marked by *wel* 'indeed' as in (7). *Hoeven* could also be substituted by a non-NPI modal verb by means of *self-correction*, like by the PPI *moeten* 'must' in (8) (Iatridou and Zeijlstra 2010, 2013; Homer 2012).

(6) *xx*   <u>*hoef*</u>   ***geen***   *suiker*.
    xx   need   no   sugar
    'xx does not need sugar.'
    (Wijnen and Bol 1993: abe21000.cha: line 24)

(7) *Nee,*   <u>*hoeft*</u>   ***wel***.
    no   needs   surely
    'No, it *is* necessary.'
    (Wijnen and Bol 1993: mat30603.cha: line 1467)

(8) *Deze*   <u>*hoe(ft)*</u>   *deze*   <u>*moet*</u>   ə   *op*.
    this   need(s)   this   should   ə   on
    'This needs, this should be here.'
    (Wijnen and Bol 1993: daa20910.cha: line 436)

At first sight, utterances such as (7) and (8) differ from the distribution of the NPI introduced in Section 1 (consider again examples (1) to (3)); however, this is only

apparent. The category of adult-like does not only include grammatically licensed *hoeven* but also all child utterances of *hoeven* that are found in adult Dutch as well. Expression (7) is counted as pragmatically licensed and falls under adult-like because Dutch adults sometimes produce such utterances under a contrastive context to deny a preceding utterance, such as *Dat hoeft niet* 'That is not necessary'. Data of adult Dutch collected in *het Corpus Gesproken Nederlands* 'the Spoken Dutch Corpus', Oostdijk 2004, hereafter the CGN) also show that Dutch speakers indeed utter such instances containing *hoeven* appearing under a contrastive focus marked by *wel* 'surely' (see Appendix 2).[9] The same line of reasoning applies to utterances such as (8).[10]

On the other hand, child utterances of *hoeven,* unlike those of adult Dutch speakers, fell under the category of non-adult-like; but because these utterances may differ from their adult-like counterparts for different reasons, we further divided children's utterances containing non-adult-like *hoeven* into three subcategories. This allowed us to examine the exact nature of the non-adult-likeness.[11] The subcategory *licensed by pseudo-licensers* refers to those utterances containing a linguistically legal licenser that was not phonologically realized in a proper adult-like way, such as in (9) in which the anaphoric negation *nee* is uttered instead of the required *geen* 'no(ne)'.[12] When a linguistically legal licenser of *hoeven* was only contextually present, by means of incorrect ellipsis as given in (10), it counted as *licensed by contextual present licenser*.[13] Finally, the subcategory of *not licensed* covered *hoeven*'s occurrences in the absence of any linguistic legal licenser and/or a proper pragmatic context as illustrated in (11).

---

9 Data of adult Dutch were collected from *Component a* "Spontaneous conversations ('face-to-face')", *Component c* "Spontaneous telephone dialogues (recorded via a switchboard)" and *Component d* "Spontaneous telephone dialogues (recorded on MD via a local interface)" of the CGN (Oostdijk 2004).

10 As one of the reviewers correctly pointed out, we were not able to judge whether *hoeven* would have been licensed in case of self-correction. The crucial point for us for analyzing self-correction as adult-like is that this is also something that Dutch adults do (see Appendix 2).

11 Other types of non-adult-likeness, such as at syntactic level (e.g., ungrammatical word orders), at morpho-syntactic level (e.g., ungrammatical verbal inflection), at lexical level (e.g., *waf waf* instead of *hond(je)* 'dog(gy)', *blaan blaan* instead of *banaan* 'banana'), are all disregarded in the current research.

12 This type of non-adult like licensing of *hoeven* is also mentioned in Van der Wal (1996). We will come back to this point after presenting Table 2.

13 Utterance (10) is analyzed as an instance of non-adult-like because such ungrammatical ellipsis was not found in adult data collected in the CGN and it was rejected by two Dutch informants.

(9) *Ik* *hoef* *pit* *in (.)* **nee**.
    I   need seed inside no
    Lit. 'I need a seed inside no.'
    'I don't need a seed inside.'
    (Wijnen and Bol 1993: mat20926.cha: line 654)

(10) *Deze hoef ik* **niet**. *deze* *hoef* *ik ook*.
    this   need I   not   this   need I   also
    'I do not need this one, and this one neither.'
    (Van Kampen 1994: sarah26.cha: line 363)

(11) *Ik* *hoef* *drinken*.
    I   need drink
    'I want to drink something.'
    (Wijnen and Bol 1993: mat30113.cha: line 515)

If, after taking the linguistic environment and available contextual information into account, it was still impossible to analyze *hoeven*'s licensing status, an instance of unclear situation was counted. An example of such a case is in (12), where an incomplete utterance is concerned.

(12) *(I)k* *hoef* ...
    I   need
    'I need ...'
    (Wijnen and Bol 1993: jos30110.cha: line 1275)

This categorization of children's use of NPI's leads to the results in Table 2.

In each developmental stage, the percentage of adult-like utterances of *hoeven* exceeded 90% of the total amount. However, in order to examine the first prediction motivated in Section 3 (the absence of overgeneralization errors), we should further analyze the category of non-adult-like utterances of *hoeven*, although the percentage of this category is less than 5% in Stage I and II, and even zero in Stage III.

In the case of licensed by pseudo-licensers, i.e., 0.85% in Stage I, zero percent in both Stage II and III, children employed head-shaking, anaphoric negation *nee* (or under-articulated *nə* or *ə*), or intonation contour to express semantic negation, although a properly phonologically realized negative marker was absent (see also Van der Wal 1996). Children's employment of a pseudo-licenser provides evidence for their awareness of *hoeven*'s sensitivity to the negative polarity of an utterance; therefore *hoeven* licensed by a pseudo-licenser cannot be taken to represent any overgeneralization errors. *Hoeven* licensed by a negation that was

only present in discourse context cannot be considered an overgeneralization error either. This is because the contextually present negation indicates children's awareness of some restriction of *hoeven* as well. The subcategory of not licensed, finally, might identify children's overgeneralization errors. However, since the percentage of this subcategory throughout the whole development of child Dutch was extremely marginal, i.e., 0.85% in Stage I, 1.23% in Stage II and zero percent in Stage III, we interpret the subcategory of not licensed to represent children's performance errors only.[14] Hence, the first prediction is confirmed: Dutch children do not overuse the NPI in non-licensing contexts in any developmental stage.

Table 2: Distribution of *hoeven* in child Dutch by licensing status

Category	Subcategory	Stage I age: 1–3	Stage II age: 3–4	Stage III age: 4–5
Adult-like	Linguistically licensed	102 (87.18%)	68 (83.95%)	49 (96.08%)
	Pragmatically licensed	5 (4.27%)	4 (4.94%)	0
	Self-correction	2 (1.71%)	1 (1.23%)	1 (1.96%)
Total adult-like		**109 (93.16%)**	**73 (90.12%)**	**50 (98.04%)**
Non-adult-like	Licensed by pseudo licensers	1 (0.85%)	0	0
	Licensed by contextually present negation	3 (2.56%)	2 (2.47%)	0
	Not licensed	1 (0.85%)	1 (1.23%)	0
Total non-adult-like		**5 (4.27%)**	**3 (3.7%)**	**0**
Unclear situations		**3 (2.56%)**	**4 (4.94%)**	**1 (1.96%)**
TOTAL		**117**	**81**	**51**

As for the second and the third prediction, we further analyze all children's adult-like utterances containing linguistically licensed *hoeven* depending on the type of its licensing contexts. Adopting a standard classification of different types of negative expressions argued to play a role in NPI licensing (Zwarts 1986, 1998; Van der Wouden 1997), we distinguish three types of licensing conditions for the

---

14 Since the subcategory of *unclear situations* refer to children's utterances in which the licensing status of *hoeven* was impossible to categorize depending on both the linguistic environment and available contextual information, this category can hardly represent children's genuine erroneous use of the NPI. However, even if we were to assume the utterances of this subcategory all contained unlicensed *hoeven*, the percentage of the subcategory of *not licensed* would still be too low to represent systematic overgeneralization errors.

target NPI. If *hoeven* occurred under the scope of the negative marker *niet* as in (13), then it counted as an instance of *sentential negation* licensing. Licensed by a negative quantifier, such as *niks* 'nothing' in (14a) and *geen* 'no(ne)' in (14b), then it was counted as an instance of *hoeven* licensed by *negative indefinites*. If a form of negation weaker than *niet* or negative indefinites licensed *hoeven*, such as *alleen* 'only' in (15), it was counted as an instance of *weaker negation* licensing. Results following from this further analysis are presented in Table 3.

(13) *Nee, ik <u>hoef</u> **niet** naar de wc.*
      no   I   need   not   to   the   toilet
      'No, I do not have to go to the bathroom.'
      (Wijnen and Bol 1993: abe30308.cha: line 928)

(14) a. *xxx <u>hoef</u> **niks** van.*
         xxx   need   nothing   from
         Lit. 'xxx needs nothing.'
         '(I) do not need any of that (chocolate milk).'
         (Van Kampen 1994: sarah43.cha: line 615)
     b. *xx <u>hoef</u> **geen** suiker.*
         xx   need   no   sugar
         Lit. 'xx needs no sugar.'
         '(I) do not need any sugar.'
         (Wijnen and Bol 1994: abe21000.cha: line 24)

(15) ***Alleen*** *<u>hoef</u> je zo te doen.*
      only   need   you   so   to   do
      'You only have to do so.'
      (Van Kampen 1994: sarah44.cha: line 271)

Table 3: Distribution of *hoeven* in child Dutch by licensing condition

Licensing condition	Stage I (age: 1–3)	Stage II (age: 3–4)	Stage III (age: 4–5)
Sentential negation (*niet* 'not')	101 (99.02%)	67 (98.53%)	42 (85.71%)
Negative indefinites (*geen* 'no(ne)', *niks* 'nothing', etc.)	1 (0.98%)	0	5 (10.2%)
weaker negation (*alleen* 'only', *weinig* 'few', etc.)	0	1 (1.47%)	2 (4.08%)
TOTAL	**102**	**68**	**49**

Whereas the distribution of *hoeven* across different types of negative contexts in Stage I does not significantly differ from its distribution in Stage II (p = .638, >α = .01, df = 2), we do find a significant difference between children younger than four years old and their older counterparts in Stage III (p = .001, <α = .01, df = 2). Moreover, we find that the contributor to this significant effect is the emergence of negative indefinites as a *new* type of licensers for *hoeven* in Stage III (Std. Residual: −1.7 and 3.2 at α = .01 before and after the age of four, respectively). This means that children in Stage III employ significantly more types of negative contexts to license the target NPI compared to children below the age of four. Therefore we confirm the second prediction that the distribution of *hoeven* is more restricted in early than in late child Dutch.
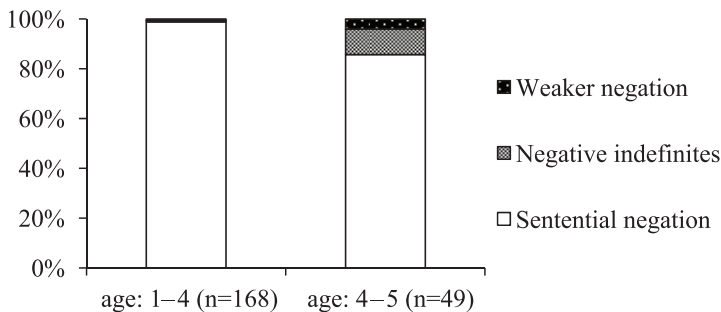
Fig. 1: Developmental pathway in the acquisition of *hoeven*

    The corpus results discussed so far can also be represented in Figure 1. As this figure shows, the acquisition of *hoeven* exhibits two distinctive developmental periods with the age of four as a watershed. Focusing first on *hoeven*'s distribution in early child Dutch, we find that this NPI is always attested under the scope of the sentential negative marker *niet*; on the other hand, in late child Dutch, the NPI is allowed in (at least) negative contexts introduced by a negative indefinite as well.[15] That *hoeven* in early child Dutch can apparently only be licensed by *niet* is exactly what we expected, since the negative marker *niet* turns out to be the most frequent licenser of *hoeven* attested in child-directed speech, occurring 80.81% of the time (see

---

15 In the current study, early child Dutch refers to the developmental period before the age of four; late child Dutch represents the period after the age of four but before the age of five since the corpus research only covers Dutch children between one and five years old.

Appendix 1).[16] Thus, the distribution of *hoeven* is restricted to negative con-texts introduced by *niet* in a child's initial analysis, but this distribution is ex-panded later on. This is completely in line with our last prediction motivated in Section 3. Given the massive co-occurrence of *hoeven* and *niet* within the same clause, Dutch children only produce utterances of the NPI under the scope of *niet* in early stages.

With all three predictions confirmed, we conclude that Dutch children indeed employ a conservative widening learning strategy in their acquisition of the tar-get NPI. This conservative strategy explains how and why children are able to detect a restricted distribution of an NPI to certain negative contexts only – even in the absence of both direct and indirect negative evidence. But how exactly does this hypothesis explain the developmental pathway illustrated in Figure 1? This question is addressed in the following section.

6 Explaining the acquisitional pathway

As shown in Section 5, the acquisition of the NPI *hoeven* exhibits two stages. In the first stage, Dutch children only allow the target NPI to appear under the scope of the negative marker *niet*. In the subsequent stage, Dutch children also utter *hoeven* in negative contexts introduced by negative indefinites, such as *geen* and *niks*. Here we take these two developmental stages in the acquisition of *hoeven* to represent two distinct analyses of this NPI by Dutch children of different ages, in particular an initial analysis A1 in early child Dutch and a reanalysis A2 in late child Dutch. Under the hypothesis of conservative widening, A1 and A2 must sat-isfy the following requirements. First, both A1 and A2 should be triggered by pos-itive evidence. Second, A1 should be more rigid than A2 due to an extending, weakening-down process in acquisition predicted by the conservative widening strategy: A1 should form a subset of the output of A2. But what are A1 and A2 and how do Dutch children establish A1 and A2 in early and late child Dutch, respectively?

According to the conservative widening learning strategy, the acquisition of *hoeven* takes place on the basis of positive evidence only. Moreover, we assume distributional information to play a crucial role in the initial stage of analyzing the NPI, as we do not assume any inherent awareness of *hoeven* being an NPI.

---

16 Data concerning *hoeven*'s distribution in child-directed speech are collected from the follow-ing subcorpora of the CHIDLES database (MacWhinney 2009): *BolKuiken* (Bol and Kuiken 1990), *Groningen* (Wijnen and Bol 1993), *VanKampen* (Van Kampen 1994) and *Wijnen* (Wijnen 1988, 1992; Elbers and Wijnen 1992). Only utterances containing linguistically licensed *hoeven* were taken into account.

Therefore, investigating the distributional pattern of *hoeven* in language input is a logical starting point to answer the questions elaborated above. As already mentioned in the previous section, clearly the most frequent licenser for the target NPI in child-directed Dutch turns out to be the negative marker *niet*, which is used 80.81% of the time (see also Appendix 1). Among these utterances, we found that the percentage of co-occurrence of *hoeven* and *niet* with a distance between zero (i.e., adjacent co-occurrence) and three syllables reached 97.32%.[17] Hence, *niet* is not only the most frequent licenser in child-directed speech, but it is also (nearly) adjacent to *hoeven*. Since children are already sensitive to this type of distributional information at a very early age (probably even before the age of two, cf. Santelmann and Jusczyk 1998), we propose that Dutch children's initial analysis A1 is that *hoeven* is lexically associated with *niet* as [HOEVEN NIET].[18]

A1 as [HOEVEN NIET] is very strict and conservative since it can only generate *hoeven*'s occurrence under the scope of *niet*; but it provides Dutch children an explanation for more than 80% of the language input. Nonetheless, there is still 20% of the input data that cannot be captured by A1 as [HOEVEN NIET]. These data may consist of *hoeven*'s occurrence under the scope of other forms of negation, for instance, under the scope of negative indefinites *geen*, *niks*, quasi-universal quantifiers *niet iedereen* 'not everybody' or exclusive adverbs *alleen*, *slechts* 'merely'. Such input data – which falsify A1 – trigger Dutch children to make a reanalysis of the target NPI. We observe that of the remaining 20% of language input containing *hoeven*, more than 15% are utterances in which this NPI is licensed by negative indefinites (Appendix 1). Thus, Dutch children are expected to establish a reanalysis that allows *hoeven* to appear not only under the scope of *niet*, but also in negative sentences introduced by negative indefinites.

Interestingly, however, negative indefinites can be analyzed to contain an abstract negation incorporated in their underlying syntactic structure. Following Jacobs' (1980) analysis of negative indefinites, originally proposed for German,

---

17 An investigation of *hoeven* licensed by the negative marker *niet* in child-directed Dutch (Wijnen 1988, 1992; Bol and Kuiken 1990; Wijnen and Bol 1993; Elbers and Wijnen 1992; Van Kampen 1994) leads to the following distribution pattern depending on the distance in terms of syllables between *hoeven* and *niet*: the percentage of adjacent co-occurrence is 42.58%; the percentage of co-occurrence with a distance of one syllable is 36.91%; the percentage of co-occurrence with a distance of two syllables is 10.73%; the percentage of the co-occurrence with a distance of three syllables is 5.68%; and the percentage of the co-occurrence with a distance greater than three syllables is 4.1% (see also Appendix 5).

18 Note that we do not want to claim that the children we studied are in the same stage as those studied by Santelmann and Jusczyk (1998). The crucial point here is that the first analysis these children made is very likely to be a lexical analysis, given the distribution of *niet* and *hoeven* in child-directed speech.

negative quantifiers represent a syntactically complex structure that consists of an abstract negation (NEG) and an existential quantifier (∃). This syntactically decompositional view on negative quantifiers is argued to hold for Dutch as well (see Rullmann 1995; Zeijlstra 2011; among others). In line with Zeijlstra (2011), negative indefinites in Dutch can be analyzed in the following syntactic way.
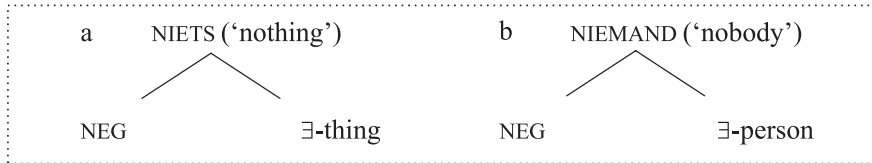
       a   NIETS ('nothing')   b   NIEMAND ('nobody')

     NEG   ∃-thing   NEG   ∃-person

Fig. 2: Decompositional analysis of negative quantifiers in Dutch

The syntactic structure of Dutch negative indefinites as proposed above is motivated by the existence of so-called split-scope readings of negative indefinites (Rullmann 1995; Penka and Zeijlstra 2005; Penka 2012; Zeijlstra 2011; Iatridou and Zeijlstra 2013). See an example in (16) below.

(16) *Je   mag   niks   eten.*
      you   may   nothing   eat
      a. 'You are allowed to eat nothing.'   may > NEG ∃-thing
      b. 'There is no specific thing that you are allowed to eat.'
                                                     NEG > ∃-thing > may
      c. 'It is not the case that you are allowed to eat a thing.'
                                                       NEG > may > ∃-thing

The example above has three readings, i.e., a narrow scope reading in which the negation together with the existential quantifier are interpreted under the scope of the modal verb *may* (16a); a wide scope reading in which the negation and the existential quantifier together scope over the modal verb *may* (16b); and a split-scope reading in which the negation and the existential quantifier are separately interpreted, such that the negation scopes over *may* and this modal verb in turn takes scope over the existential quantifier (16c). To illustrate the difference between these three possible interpretations of a Dutch sentence like (16), let us first consider a context in which a child would like to have some cookies or candies right before going to bed and asks his/her mum whether he/she may eat something. Suppose the mother utters (16) in response. A narrow scope interpretation of (16) would then mean that the mother gives her child the permission or authorization to eat nothing. A wide scope interpretation of the negative indefinite *niks*, on the other hand, generates a scenario in which the mother does not have in

mind any specific thing that the child is allowed to eat right before going to bed. Finally, by assigning a split-scope interpretation of the mother's utterance (16), the child reaches a situation in which he/she is just not allowed to eat anything before sleeping. In the pragmatic context we are describing here, the split-scope interpretation (16c) is by far the most salient and the only one possible. The availability of this split-scope reading by both the mother and the child in turn provides empirical evidence for the syntactic structure of Dutch negative indefinites as shown in Figure 2 in both adult and child language (cf. Zeijlstra 2011).

　　Given this decompositional analysis of negative quantifiers, we propose that in late child Dutch, *hoeven* must be lexically associated with an abstract negation that is present in all negative indefinites. This abstract negation is referred to as NEG; consequently, A2 can therefore be represented as [HOEVEN NEG].[19] Compared to the constructional A1 as [HOEVEN NIET] in early child Dutch, A2 as [HOEVEN NEG] is a more abstract and general analysis since it not only generates *hoeven* under the scope of the negative marker *niet*, but also in all the other contexts containing an abstract negation. This abstract NEG can be phonologically realized by the specific negative marker *niet*, resulting in licensing of *hoeven* by the negative marker, or spelled-out as the negation incorporated in negative indefinites, leading to *hoeven*'s occurrence under the scope of negative indefinites. This is exactly how A2, i.e., the reanalysis of the initial A1, provides an explanation for the wider distribution of the target NPI in all negative contexts in late child Dutch. The developmental process from the lexical constructional analysis as [HOEVEN NIET] to the abstract analysis as [HOEVEN NEG] in child Dutch completely obeys the conservative widening learning strategy. Note that both the analyzing and the reanalyzing processes are triggered by positive evidence only. At the same time, since A2 is established in such a way that the set of its output takes the set of the output generated by A1 as its subset, as logico-semantically the set of negative environments introduced by *niet* form the subset of the set of negative contexts introduced by a negative indefinites,[20] the whole acquisitional process of the target NPI can be described best in terms of the Subset Principle. More importantly,

19 Here we assume that the process of generalizing NEG from the sentential negative marker *niet* and negative indefinites such as *niks* or *niemand* is made possible by the following ingredients: a general learning strategy of abstracting structures or properties from concrete and lexical input and children's knowledge of the presence of an abstract negation NEG incorporated in Dutch negative indefinites (see our example of the child eating before bedtime). Due to space limitations, we do not discuss this generalizing step in detail but leave this for further exploration (Lin et al. in prep.).

20 The negative marker *niet* is an *anti-morphic* function. A function *f* is anti-morphic iff for every arbitrary X and Y, it holds: $f(X \cap Y) \Leftrightarrow f(X) \cup f(Y)$ and $f(X \cup Y) \Leftrightarrow f(X) \cap f(Y)$ (adapted from Van der Wouden 1997). Negative indefinites such as *niks* 'nothing' or *geen* 'no(ne)' are *anti-additive*.

the developmental pathway motivated here signifies why frequency pattern in the input alone is not sufficient in explaining the acquisition of the NPI: without the linguistic knowledge of the decomposability of negative quantifiers at the syntactic level, Dutch children would not be able to extract the abstract negation NEG to develop an abstract representation of the target NPI from the lexical construction. The question is: how do we know that such linguistic knowledge is already available to children younger than the age of four?

In order to examine to what extent the decompositional analysis of negative indefinites is acquired in the first stage, we investigated spontaneous speech data of younger Dutch children. We found that as early as three years old, children are already able to produce utterances in the contexts where a split-scope interpretation of a negative quantifier is present. Altogether, 52 such utterances were attested in early child Dutch; two examples are illustrated below. This means that before the age of four, Dutch children have already acquired the syntactic structure of negative indefinites as elaborated in Figure 2. The reanalyzing process of *hoeven* at approximately the age of four is therefore not a speculation but supported by independent evidence.

(17) *Je   mag   geen   deurtje   lenen.* (2:11;09)
you   may   no   door   borrow
'You are not allowed to borrow a door.' NEG > may > ∃-door
(Wijnen and Bol 1993: jos21109.cha: line 356)

(18) *Ik   kan   niks   horen.* (3:06;21)
I   can   nothing   hear
'I am not able to hear a thing.' NEG > can > ∃-thing
(Wijnen 1988: 30621.cha: line 1133)

The reanalysis of [HOEVEN NEG] can definitely explain an extremely large amount of children's input. However, at first sight, this abstract A2 does not appear to account for all input data Dutch children receive. This is because in a small amount of the input, i.e. 4.04% (Appendix 1), the target NPI is licensed in negative contexts that are introduced by other operators than the negative marker *niet*

A function f is anti-additive iff for every arbitrary X and Y, it holds: $f(X \cup Y) \Leftrightarrow f(X) \cap f(Y)$ (adapted from Van der Wouden 1997).

If a function f is anti-morphic, then it is also always anti-additive, but not the other way around, since if a function f satisfies $f(X \cup Y) \Leftrightarrow f(X) \cap f(Y)$ and $f(X \cap Y) \Leftrightarrow f(X) \cup f(Y)$, which means that it is anti-morphic, then it always satisfies $f(X \cup Y) \Leftrightarrow f(X) \cap f(Y)$, i.e., being anti-additive. Thus, the set of anti-morphic contexts and that of anti-additive contexts stand in a subset relationship with each other.

and negative indefinites. Although we do not yet observe any significant use of *hoeven* in such contexts in late child Dutch, as can be seen from Table 3, our account actually predicts that Dutch children do not need to further analyze the target NPI even when they are confronted with *hoeven*'s occurrence in weaker kinds of negative contexts than those introduced by *niet* or a negative indefinite. This is because all licensing contexts for *hoeven* observed in adult Dutch already follow from children's abstract reanalysis of this NPI in the later stage. As already mentioned, A2 [ʜᴏᴇᴠᴇɴ ɴᴇɢ] allows *hoeven* to occur in all linguistic contexts that contain a decomposable abstract negation ɴᴇɢ. In the residual of the input data, we found that all those weaker types of negative environments also contain a lexically decomposable ɴᴇɢ, because all contexts introduced by semi-negative expressions, such as *weinig* 'few' *niet iedereen* 'not everybody', or exclusive adverbs, like *alleen*, *slechts* 'merely', contain underlying abstract negations as well (Von Fintel and Iatridou 2003; Penka 2011; Iatridou and Zeijlstra 2013).

Given this line of reasoning, the fact that *hoeven* is still not attested under the scope of these weaker forms of negative expressions after children's establishment of the abstract A2 as [ʜᴏᴇᴠᴇɴ ɴᴇɢ] can only be explained as follows.[21] Until the age of five, Dutch children do not show sufficient evidence of having acquired the lexical knowledge pertaining to the negative expressions in question. For example, merely two utterances of *niet iedereen* are observed in the child Dutch data collected from CHILDES and the total amount of the utterances containing *weinig* attested in child language is only nine (see Appendix 4). Thus, the absence of *hoeven* in those weaker negative contexts can be explained by children's lack of the relevant lexical knowledge. On the other hand, although Dutch children of both age groups are able to productively use *alleen* and *pas* 'not until' – a temporal version of the exclusive *alleen* – in their spontaneous speech (see Appendix 4), the non-significant occurrence of *hoeven* under the scope of these two operators may still result from the complexity of the operators as such. Pragmatic knowledge of scalar implicatures and presuppositions is essential for a child to analyze *alleen* as exhibiting a decomposable and abstract negation. With respect to *pas*, a temporal quantifier, children are required to acquire temporal concepts and their grammatical realizations before they may be able to extract the abstract negation incorporated in *pas*. The non-occurrence of the NPI *hoeven* in such weaker types

---

21 Of course, what children produce in their spontaneous speech does not always equal what they *can* produce. Because of the limitation of the methodology of the current research, i.e., a corpus investigation, we are not able to exclude another possibility: children are already able to license *hoeven* by different weaker forms of negation after their reanalyzing process, but they just do not (yet) spontaneously produce such utterances. We leave this exploration for further research (Lin et al. in prep.).

of negative environments in late child Dutch is therefore not a consequence of how children analyze the NPI, but how they acquire and analyze weaker negative expressions such as *weinig*, *slechts*, *niet iedereen*. This leads to the conclusion that A2 containing an abstract negation as presented as [HOEVEN NEG] is the final analysis of *hoeven* at the end of language acquisition, i.e., the adult analysis.

The adult analysis as [HOEVEN NEG] makes a prediction for how the NPI is distributed in adult Dutch, namely that *hoeven* is restricted to only those negative contexts containing a (decomposable) negation. An investigation of a total of 1694 utterances containing the target NPI collected in the CGN confirms this prediction (see a similar result observed in written Dutch in a previous corpus investigation by Hoeksema 1997: 11: Figure 5).[22] As is shown below, *hoeven* is allowed to appear either in a negative context introduced by *niet* in (19) or under the scope of a negative indefinite as in (20). It may also appear in weaker kinds of negative contexts as in (21) and (22), since the abstract negation NEG present in these contexts can license the NPI, satisfying the reanalysis [HOEVEN NEG]. However, the adult analysis as [HOEVEN NEG] bans *hoeven* from those contexts in which an abstract negation is absent as in (23) to (26).

(19)   *Jan*   *hoeft*   *vandaag*   **niet**   *te*   *werken.*
     John   needs   today   NEG   to   work
     'John does not have to work today.'

(20)   **Niemand**   *hoeft*   *vandaag*   *te*   *werken.*
     NEG-body   needs   today   to   work
     'Nobody has to work today.'

(21)   **Niet iedereen**   *hoeft*   *vandaag*   *te*   *werken.*
     NEG   everybody   needs   today   to   work
     'Not everybody needs to work today.'

(22)   **Alleen**   *Jan*   *hoeft*   *vandaag*   *te*   *werken.*[23]
     NEG other than   John   needs   today   to   work
     'Only John has to work today.'

(23)   **Iedereen*   *die*   *hoeft*   *te*   *werken*   *moet*   *nu*   *weg.*
     everybody   COMP   needs   to   work   must   now   away
     Intended: 'Everybody that has to work must leave now.'

---

22 See footnote 9.

23 We adopt a decompositional analysis of exclusive adverbs as proposed in Von Fintel and Iatridou (2003), among others. According to this analysis, exclusive adverbs such as English *only* and Dutch *alleen* contain a negation (NEG) and an exceptive component *other than*.

(24) **Als   Jan   <u>hoeft</u>   te   rijden   moet   hij   mij   ophalen.*
     if   John   needs   to   ride   should   he   me   pick up
     Intended: 'If John has to ride he should pick me up.'

(25) **Wat   <u>hoeft</u>   Jan   vandaag   te   doen?*
     what   needs   John   today   to   do
     Intended: 'What does John have to do today?'

(26) **Jan   <u>hoefde</u>   gisteren   te   werken.*
     John   needed   yesterday   to   work
     Intended: 'John had to work yesterday.'

Given the examples above, it is obvious that the adult analysis as [ʜᴏᴇᴠᴇɴ ɴᴇɢ] makes *hoeven* exhibit a wider distribution than Dutch *ook maar* 'at all', restricted to anti-additive contexts only (Van der Wouden 1997; Giannakidou 1997),[24] but a narrower distribution than English *any*, allowed in all downward entailing contexts (cf. Ladusaw 1979).[25] As mentioned in Section 1, NPIs come about in different strengths, depending on the set of negative contexts that may license them. We therefore conclude that *hoeven* is a so-called strong/weak NPI that is in between *ook maar* and *any*, a typical strong and a typical weak NPI, respectively.[26]

    Moreover, the adult analysis as [ʜᴏᴇᴠᴇɴ ɴᴇɢ] suggests that *hoeven* is an NPI due to its lexical connection with an abstract negation ɴᴇɢ in a syntactic way: ɴᴇɢ is rooted in *hoeven*'s lexical representation but can be realized elsewhere in the sentence, for instance, as a negative marker *niet* or as the negation incorporated in a negative quantifier. This analysis of *hoeven* mirrors Postal's (2000) treatment of NPIs – an approach to properties underlying NPI-hood. However, this does not automatically mean that all NPIs should and can be identified by Postal (2000). In fact, we only expect NPIs of the same strength as *hoeven*, i.e.,

---

24 A function *f* is anti-additive iff for every arbitrary *X* and *Y*, it holds: $f(X \cup Y) \Leftrightarrow f(X) \cap f(Y)$ (adapted from Van der Wouden 1997).

25 A function *f* is *downward entailing* iff for every arbitrary *X* and *Y* it holds: $f(X \cup Y) \Rightarrow f(X)$ and/or $f(X \cup Y) \Rightarrow f(Y)$ (adapted from Van der Wouden 1997).

26 Depending on in exactly which type of negative contexts an NPI can be licensed, four NPI-strengths are distinguished in the literature: superstrong NPIs that are allowed to appear under the scope of sentential negation only, such as Dutch *mals* 'mild'; strong NPIs that survive in all anti-additive contexts, like Dutch *ook maar* 'at all'; weak NPIs that are fine in all downward entailing contexts as is observed for English *any*-terms; and superweak NPIs such as Mandarin *shenme* 'a (thing)' are licensed in all non-veridical contexts, which cannot entail the truth of an embedded proposition (definition adapted from Zwarts 1993). For related discussion the reader is referred to Zwarts (1986, 1998), Nam (1994), Van der Wouden (1997), Giannakidou (1998, 2002), Hoeksema (2000), Lin (2011a, 2011b).

strong/weak NPIs, to bear a lexical connection with an abstract negation in their syntactic representation as [NPI NEG]. Likewise, since previous analyses to disentangle the property underlying the NPI-hood are proposed for the minimizing indefinite NPIs, denoting low-scale endpoints (such as English *any*, see Kadmon and Landman 1993; Krifka 1994; Chierchia 2004, 2006), we cannot expect *hoeven* – a universal NPI expressing necessity (cf. Kratzer 1981, 1991) – to be explained in a similar way; *hoeven*, given its universal force, requires a separate analysis for its NPI status.

This means that our investigation of the acquisition of the Dutch NPI also sheds light on a theoretical question of whether NPIs form a natural category in human languages that share a similar base underlying their polarity sensitivity. Because the acquisitional pathway from a lexical construction of [HOEVEN NIET] to a non–construction-specific configuration of [HOEVEN NEG] clearly shows children's development of a lexical dependency between *hoeven* and an abstract negation NEG, we conclude that *hoeven* has become an NPI due to its lexical association with the abstract negation NEG. The reason underlying its NPI-hood is therefore a direct result of children's acquisitional trajectory.

7 Alternative approaches

This section discusses two alternative approaches to the attested pathway of *hoeven*. The first approach is that of Van der Wal (1996) that takes children's initial step in acquiring the NPI as a consequence of their awareness of *hoeven*'s NPI status from the acquisitional onset and argues the developmental pathway of *hoeven* stems from the acquisition of different negative expressions in Dutch (Van der Wal 1996). The second approach is developed from the distributional learning mechanism of Mintz (2002) and Mintz et al. (1995, 2002) (see also Cartwright and Brent 1997; Redington et al. 1998), and we refer to this as *a lexical approach*. This approach would expect children to add up different lexical constructions of *hoeven* in later stages as well, instead of establishing an abstract reanalysis as [HOEVEN NEG]. However, this section shows that both alternative explanations are problematic in explaining the attested acquisitional pathway.

7.1 Van der Wal (1996)

Van der Wal (1996) is the first to address the learnability problem raised by the existence of NPIs. Postulating the hypothesis of conservative widening, she aimed to unravel the problem by investigating Dutch children's acquisition of two polarity items *hoeven* and *meer* 'anymore' via both corpus research and

experiments; however, no exclusive evidence was claimed to be found for children's employment of this learning strategy in the acquisition of the NPIs. Instead, she argued for an acquisitional pathway of the two NPIs that stems from how and when Dutch children acquire negative expressions. Because "from the onset [Dutch children] are aware of the fact that NPIs cannot appear in all configurations" (Van der Wal 1996: 179), they only allow *hoeven* (and also *meer* 'anymore') to appear under the scope of a negation. Since *niet* is the only negation acquired in early child Dutch, the target NPIs are always licensed by this single licenser in an initial stage. In a later developmental stage, when children have acquired different negative expressions in Dutch, for instance, negative indefinites *geen* and *niemand*, they start to license the NPIs by these late-acquired negations as well. This implies that *hoeven*'s restricted distribution to under the scope of the negative marker *niet* only in early child Dutch is a consequence of children's limited knowledge of what counts as a negation in adult grammar. This alternative explanation takes the acquisition of the NPI *hoeven* to be in tandem with that of Dutch negation and the observed widening pattern therefore represents children's (in)complete knowledge of various negative expressions at different stages (see also Koster and Van der Wal 1996). Although she speculated that "from the onset in children's speech [NPIs] have a *principled* restricted distribution, rooted in knowledge of the essence in NPI licensing" (Van der Wal 1996: 179), she did not discuss how and why Dutch children would have access to such knowledge. Thus, her speculation of children's awareness of *hoeven* being an NPI – already available from the very onset of language acquisition – leads to conjecturing some innate universal principle in NPI use for not only Dutch-acquiring children but also their English and German counterparts (Van der Wal 1996: Ch. 5: 5.3). Our explanation in Section 6 clearly shows that the distributional properties of *hoeven* in child-directed Dutch alone are already sufficient for children to make the first step in its acquisition; it is therefore unnecessary to adopt innate knowledge here. Moreover, in what follows, we present three pieces of evidence against *hoeven*'s developmental pathway determined by children's acquisition of negation as proposed by Van der Wal.

First, we found in our corpus data that Dutch children younger than the age of four are already able to spontaneously and productively use at least three of the five negative indefinites attested as licensers for *hoeven* in adult Dutch, i.e., *geen*, *niks* (*niets*) and *niemand*. The frequency data presented in the table below reject Van der Wal's proposal.[27] Therefore, that *hoeven's* distribution is restricted

---

27 Data were collected from the same subcorpora in the CHILDES database (MacWhinney 2009) as introduced in Section 4.

to under the scope of *niet* only in early child Dutch does not lie in younger children's lack of lexical knowledge of diverse negative indefinites. This means that the developmental pattern observed in the acquisition of *hoeven* (see Figure 1) cannot be taken as a reflection of Dutch children's (in)complete acquisition of various negative operators of different ages.

Table 4[28,29]: Frequency of different negative indefinites in early child Dutch

Negative indefinites	age: <4
geen 'no(ne)'	285
niks (*niks*) 'nothing'	202
niemand 'nobody'	23
nooit 'never'	9
nergens 'nowhere'	0
TOTAL	**519**

Secondly, we observed that in early child Dutch, negative indefinites already occur in combination with a variety of modal verbs, such as *willen* 'will', *kunnen* 'can' and *mogen* 'may', which are polarity insensitive modal verbs, and *moeten* 'must', which counts as a PPI in the adult grammar (Iatridou and Zeijlstra 2010, 2013; Homer 2012). Altogether, 51 instances were attested in this respect. The co-occurrence of negative indefinites and almost every other modal verb with the exception of *hoeven* in early child Dutch necessarily excludes the explanation of Van der Wal (1996). If the distribution of *hoeven* in early child Dutch were determined by children's incomplete acquisition of negation, we would not be able to account for why, in the first stage, only *hoeven* is not allowed to appear in negative environments introduced by a negative indefinite.

Finally, we found overgeneralization errors of the negative marker *niet* to the negative quantifier *geen* when it comes to the licensing of *hoeven*. Such overgeneralization errors concern ungrammatical utterances of Dutch children younger than the age of four. See the a-sentences below, in which an asterisk denotes their ungrammaticality in adult grammar; the grammatical counterparts in adult Dutch are given in the b-sentences.

---

28 See also footnote 11. Only the frequency of the five negative indefinites that are attested in adult Dutch was included here (see Appendix 3). Note that two instances of concord-like use of negative indefinites are excluded, such as *ik hoef nooit geen huisdier te hebben* (lit. 'I need never no pet to have', meaning 'I would never like to have a pet').
29 Among all the occurrences of the negative operators in Table 4, only one instance of *hoeven*-licensing is attested, in which the negative quantifier *geen* is employed.

(27) a. **Ik hoe(f) melkje **niet**.*
     I need milk not
     Intended: 'I do not want any milk.'
     (Van Kampen 1994: laura12.cha: line 34)
  b. *Ik hoef **geen** melk.*
     I need no milk
     'I do not want any milk.'

(28) a. **(Ik) hoef **niet** suiker.*
     I need not sugar
     Intended: 'I do not want any sugar.'
     (Wijnen and Bol 1993: jos300020.cha: line 102)
  b. *Ik hoef **geen** suiker.*
     I need no sugar
     'I do not want any sugar.'

Both a-sentences are not ungrammatical due to illegal licensing of *hoeven*. These utterances are ungrammatical because adult Dutch speakers would employ *geen* instead of *niet* to quantify a noun phrase, i.e., *melk* 'milk' in (27) and *suiker* 'sugar' in (28). A closer look at Dutch children's spontaneous speech production leads to the observation that children only use *niet* instead of *geen* in contexts in which *hoeven* needs to be licensed. Among a total of 168 utterances in which *hoeven* occurs under the scope of *niet* (see Table 3), we found 22 instances of this kind of overgeneralization error. Since such overgeneralizations are only restricted to the NPI, the explanation for (27a) and (28a) can hardly lie in children's incomplete acquisition of negation in the first stage. Given our central claim that the developmental pattern of the acquisition of *hoeven* represents different analyses of this NPI by Dutch children of different ages, we come to the conclusion that (27a) and (28a) can only be explained in one way: in our view, the target NPI has not yet been reanalyzed as [HOEVEN NEG] in early child language. With their strict A1 as [HOEVEN NIET], children younger than the age of four are only able to allow *niet* to license the target NPI, resulting in overgeneralization of *niet* for *geen* as shown in (27a) and (28a).

    Moreover, we attested a significant decrease of the total amount of such overgeneralizations in late child Dutch, i.e., when *hoeven* is analyzed as [HOEVEN NEG] (p = .042, df = 1). Since even before the age of four, Dutch children are already able to productively use *geen* outside the contexts of the licensing of *hoeven*, as illustrated in Table 4, the significant decrease exclusively supports our proposal that different analyses of the target NPI shape its acquisitional pathway. Therefore, the acquisition of the NPI *hoeven* does not stem from how Dutch children acquire different negative expressions in adult language.

7.2 A lexical approach

As shown in Mintz (2002), Mintz et al. (1995, 2002) and Redington et al. (1998), a distribution-based learning strategy can perfectly well predict early processes of learning grammatical categories, such as noun and verbs. Since we do not assume innate linguistic knowledge of *hoeven* being an NPI, we adopted the distributional perspective in language acquisition to explain how Dutch children initially analyze the target NPI as having a lexical dependency with the negative marker *niet* and establish a constructional analysis as [HOEVEN NIET]. When confronted with input data falsifying the lexical frame, children abandon the lexical analysis developed via the distribution-based learning modal and switch over to a more abstract but less rigid reanalysis as [HOEVEN NEG]. This step in abstracting the analysis of the target NPI is supported by the observation that children have already acquired the decomposable syntactic structure of Dutch negative indefinites before the age of four. Because the reanalysis is of great generalizability in the sense that it explains all input data, including the co-occurrence of *hoeven* and *niet*, children do not need to further analyze the NPI and [HOEVEN NEG] turns out to be the adult analysis of the NPI. The whole acquisitional process from a purely lexical analysis as [HOEVEN NIET] to a more abstract and generalizable reanalysis as [HOEVEN NEG] is captured by the Subset Principle, since the set of output of the lexical constructional analysis in the early stage forms a subset of the set of the output generated by the abstract reanalysis in the late stage.

However, one might argue that children could also continue to employ a distributional mechanism in a later stage when they receive or can analyze more input data conflicting with the lexical frame as [HOEVEN NIET]. Among these input data, we found that more than 70% exhibit a co-occurrence of *hoeven* and *geen* within a distance of three syllables (see Appendix 5). From a distributional perspective, such input evidence could still trigger Dutch children to establish a lexical dependency between *hoeven* and *geen* and construct a second lexical frame as [HOEVEN GEEN]. Along this line of reasoning, the distributional information of *hoeven* in child-directed speech (see Appendix 5) would lead Dutch children to continually establish a third lexical frame as [HOEVEN NIKS], a fourth construction as [HOEVEN MAAR], a fifth construction as [HOEVEN ALLEEN], and a sixth frame as [HOEVEN NOOIT]. As such, the acquisition of *hoeven* by older children can be described as a process of adding up lexical constructions according to the distributional properties of the NPI in the language input. This acquisitional process from a distributional perspective could also be captured by the Subset Principle due to the subset relationship between the singleton set containing a single lexical frame as [HOEVEN NIET] and the set containing multiple lexical constructions, such as [HOEVEN NIET], [HOEVEN NIKS], and [HOEVEN MAAR]. But is this

lexical explanation plausible? In other words, to what extent is the abstraction in the reanalysis as shown in Section 6 required? Our discussion of the lexical approach begins with revisiting input evidence in child-directed Dutch and the target distribution of *hoeven* in adult Dutch.

In child-directed Dutch, we found a total of 40 child-directed utterances containing *hoeven* co-occurring with *geen* within a distance of three syllables (see Appendix 6). Given that our child-directed data were collected from a total of 331.5 hours of recording (see Appendix 7), equal to approximately 33 days (based on the assumption that a child younger than the age of five is awake for 10 hours per day), this amounts to one utterance of *hoeven* under the scope of *geen* per day. Hearing *hoeven* co-occurring with *geen* within a distance of three syllables once a day does not appear to us as sufficiently frequent for children to establish a second lexical frame as [HOEVEN GEEN]. Nevertheless, we can still imagine that children might be able to form this second construction after repeatedly being confronted with the co-occurrence of *hoeven* and *geen* once per day for, let's say, two weeks.

However, if we focus on the frequency of the co-occurrence of *hoeven* with *niks* (*niets*) or *alleen* within a distance of three syllables among all input utterances containing the target NPI, it becomes particularly doubtful that Dutch children would still be able to employ the distributional learning model. The frequency of *hoeven* co-occurring with *niks* (*niets*), *maar*, and *alleen* in child-directed Dutch is 13, 7, and 4, respectively (see Appendix 6). Based on a total of approximately 33 days of data recording, these input frequencies entail merely one utterance of *hoeven* with *niks* (*niets*) every three days, one utterance of *hoeven* with *maar* every five days, and one utterance of *hoeven* with *alleen* about every nine days. This gives an idea of how limited one utterance of *hoeven* co-occurring with these licensers is and makes it improbable that children only rely on distributional information to acquire all the licensers separately.

Moreover, the lexical approach wrongly predicts Dutch children's acquisition of the NPI to be incomplete at the age of five. Due to non-robust and insignificant co-occurrence of the NPI with various non-*niet* licensers in child-directed speech, the distributional perspective would predict the acquisition of *hoeven* to be strongly successive and individually distinct because of subtle differences in distributional patterns of the NPI by individual speakers. This would, to a large extent, imply the existence of variations among adult Dutch speakers with respect to *hoeven* licensing; the distribution of the NPI restricted to a unified set of downward entailing contexts observed in adult Dutch via the CGN (Oostdijk 2004, see Section 6) would then merely be coincidental. On the contrary, our approach to the licensing of *hoeven,* in which older Dutch children establish the abstract reanalysis as [HOEVEN NEG], does not easily allow individual variation; the fact

that in adult Dutch, *hoeven* may only appear in a unified set of negative contexts containing a decomposable negation is then not attributed to coincidence either. Due to the presence of the abstract semantic negation (NEG) in their underlying analysis of the NPI, Dutch native speakers restrict *hoeven* to only those contexts in which NEG incorporated in *hoeven*'s syntactic representation can be spelled-out. However, before language learners can allow all possible licensers to license the NPI, they are required to first obtain the knowledge that these licensers contain an underlying, decomposable negation NEG (see also the related discussion in Section 6). Our approach therefore expects developmental variation, depending on how and why a leaner acquires the decomposable analysis of the licensers involved. Nevertheless, equipped with a single analysis as [HOEVEN NEG], Dutch children may potentially show adult-like distribution of the NPI. In comparison to having to collect a minimum of five lexical frames based on non-robust distributional information before achieving adult-like behavior of *hoeven*, with an added risk of large individual variations, establishing an abstract and generalizable analysis of this NPI is more economical and efficient for a child acquiring Dutch.

The lexical approach to the acquisition of the NPI is problematic from a theoretical perspective as well. Due to its ignorance of the existence of abstraction in language (acquisition), the lexical approach does not have any implications for why NPIs exhibit sensitivity to the negativity of an utterance, i.e., properties underlying NPI-hood, and why this sensitivity come about in different degrees, i.e., explanations for NPI-strength. The list of lexical frames of *hoeven* that older Dutch children might have established via the distributional learning approach does not explain why *hoeven* is an NPI; and *hoeven*'s strength as an NPI appears to us as randomly determined in this view. More importantly, the lexical approach disregards the existence of well-established semantic categories in natural languages. By considering only the distributional pattern of *hoeven* in the language input, the lexical approach to the attested acquisitional pathway overlooks the similar logico-semantic behavior of, for instance, *niks* and *nooit*, two anti-additive functions in Dutch based on a standard categorization in this respect (Zwarts 1986, 1998; Van der Wouden 1997; Giannakidou 1998). Because of a large difference between the co-occurrence frequency of *hoeven* with *niks* on the one hand, and *hoeven* with *nooit* on the other, the distribution-based explanation would predict Dutch children to construct the lexical frame [HOEVEN NIKS] much earlier in their acquisition than [HOEVEN NOOIT], leading to a disconnection between these two negative operators despite their common logico-semantic properties.

However, falsifying the distributional proposal of diverse lexical frames containing *hoeven* in late child Dutch calls for more experimental research that examines older children's performance on *hoeven* co-occurring with different

licensers of extremely low or even zero frequency in child-directed speech, such as *alleen*, *niemand* and *weinig*. This alternative approach would then predict children's bad performance when they are confronted with *hoeven* co-occurring with these licensers. Conversely, our approach that older Dutch children have already established the abstract analysis [HOEVEN NEG] would predict children's good performance independent of the input frequency of different *hoeven*-licensers – on a single condition that Dutch children have already acquired these licensers as containing an underlying, decomposable negation NEG (see also the related discussion in Section 6). We leave this experimental exploration for further research (Lin et al. in prep.).

8 Conclusions

By focusing on Dutch children's acquisition of *hoeven*, the current paper explores a solution to a learnability problem of NPIs due to the absence of both direct and indirect negative evidence (Section 2). We started by showing that a statistical learning mechanism based on statistical pre-emption cannot explain how children would be able to acquire NPIs, because lexical elements exhibiting a complementary distribution with NPIs such as *hoeven* barely exist, violating a crucial condition for a statistical mechanism to function (Section 2). We then elaborated, based on the corpus data collected in the CHILDES database, on why only the conservative widening learning hypothesis can solve the learnability problem (Section 3 and Section 4). We showed that Dutch children conservatively weaken down their analysis of the NPI in accordance with the positive evidence in their input, resulting a developmental pattern in compliance with the Subset Principle (Section 5). This prevents any type of impossible acquisitional pattern: since children cannot unlearn to overuse *hoeven* in non-licensing contexts, their acquisition of this NPI cannot start out with this kind of error, but must start out conservatively. Our investigation of *hoeven*'s distribution in child Dutch development yields an acquisitional pathway that represents two distinct analyses of the NPI by Dutch children of different ages: a constructional analysis as [HOEVEN NIET] and an abstract reanalysis as [HOEVEN NEG] in early and late child Dutch respectively (Section 6). Our approach to this pathway does not presume innate linguistic knowledge of *hoeven* being an NPI but employs distributional information in the language input as much as possible in understanding children's first attempt at analyzing the NPI. Moreover, since the syntactically decompositional analysis of Dutch negative indefinites is necessary for the process of reanalysis, we concluded that *hoeven*'s development in child Dutch is explained when both input evidence and the acquisition of independent linguistic knowledge of negative

indefinites are taken into consideration. Further corpus data collected in adult Dutch from the CGN confirmed that the reanalysis as [HOEVEN NEG] signifies exactly how adult speakers analyze the NPI, suggesting that no further analyzing process is necessary for children acquiring Dutch once they establish the abstract reanalysis as [HOEVEN NEG], which occurs after the age of four.

This paper adopts a distributional learning approach in establishing the initial lexical analysis of the target NPI, and therefore shows that it is not necessary to presume innate linguistic knowledge of *hoeven* being an NPI, contrary to Van der Wal (1996). On top of that, we also presented three pieces of evidence from child Dutch against Van der Wal's approach, which takes the developmental pattern of *hoeven* to be determined by the acquisition of Dutch negation (Section 7.1). On the other hand, however, an explanation of the attested pathway that relied solely on distributional information throughout the whole course of language development also appeared to be problematic in the particular case of the NPI (Section 7.2). This explanation makes incorrect predictions for the distribution of *hoeven* in both child Dutch and adult Dutch. Because this approach is ignorant of the existence of abstraction in (the acquisition of) different grammatical aspects, it does not contribute to previous theories of NPI-hood and NPI-strength. However, in order to examine the distribution-based learning model in late child Dutch, experimental data must be investigated. Therefore, further research on children's performance on utterances containing *hoeven* under an experimental design is in order (Lin et al. in prep.).

The current research on Dutch children's acquisition of *hoeven* leads to a prediction for the acquisition of NPIs in general. Since the conservative widening learning strategy is shown to be the only possibility to solve the learnability problem of NPIs, which is caused by the absence of both direct and indirect negative evidence, we expect all NPIs, irrespective of their strength or language, to be acquired via this learning model. Nevertheless, as illustrated in this paper, language input plays a crucial role in triggering an initial analysis and a reanalysis of an NPI; different NPIs can therefore exhibit distinct developmental pathways – both intra- and cross-linguistically – depending on their distributional properties in child-directed speech. The acquisition of NPIs thus takes place on an individual basis: acquiring some particular NPI does not signal to the learner that there is a single natural class of NPIs.

Appendices

Appendix 1: *Hoeven*'s distribution in child-directed Dutch (CHILDES) by licensing condition

Licensing condition	Count (percentage)
Sentential negation (*niet* 'not')	299 (80.81%)
Negative indefinites (*geen* 'no(ne)', *niks* 'nothing', etc.)	57 (15.41%)
Weaker negation (*alleen* 'only', *weinig* 'few', etc.)	14 (3.78%)
TOTAL	370

Appendix 2: *Hoeven*'s distribution in adult Dutch (CGN) by licensing condition

Conditions	Count (percentage)
Licensed by *niet* 'not'	1274 (76.47%)
Licensed by a negative indefinite, i.e., *geen* 'no(ne)', *niks* 'nothing', etc.	227 (13.63%)
Licensed by weaker negation, i.e., *alleen* 'only', *weinig* 'few', etc.	164 (9.845%)
In contrastive contexts	13 (1.18%)
By means of self-correction	8 (0.47%)
TOTAL	1666

Appendix 3: Frequency of *hoeven* licensed by different negative indefinites in adult Dutch (CGN)

Negative indefinites	Count (percentage)
geen 'no(ne)'	114 (51.35%)
niks (*niets*) 'nothing'	73 (32.88%)
niemand 'nobody'	8 (3.6%)
nooit 'never'	18 (8.1%)
nergens 'nowhere'	9 (4.05%)
TOTAL	222

Appendix 4: Frequency of different weaker negative expressions in child Dutch development

Weaker negative expressions	Age: <4	Age: 4–5
weinig 'few'	3	3
niet iedereen 'not everybody'	0	2
niet alles 'not everything'	1	1
slechts 'merely'	0	0
alleen 'only'	159	104
pas 'not until'	120	22

Appendix 5: Distributional properties of co-occurrences of *hoeven* and various licensers in terms of distance in syllables in child-directed Dutch (CHILDES)

Licenser	Distance in syllables	Count
niet	0	122
'not'	1	117
	2	34
	3	18
	>3	8
geen	0	18
'no(ne)'	1	19
	2	1
	3	2
	>3	0
niks (*niets*)	0	5
'nothing'	1	4
	2	2
	3	0
	>3	2
maar	0	2
'only'	1	5
	2	0
	3	0
	>3	0
alleen	0	3
'only'	1	1
	2	0
	3	0
	>3	0

Licenser	Distance in syllables	Count
nooit	0	0
'never'	1	1
	2	1
	3	0
	>3	2
Other licensers	–	3
TOTAL	–	370

Appendix 6: Co-occurrence frequency of *hoeven* with various licensers within a distance of 3 syllables in child-directed Dutch (CHILDES)

Licenser	Count	Percentage
niet 'not'	291	78.65%
geen 'no(ne)'	40	10.81%
niks (*niets*) 'nothing'	11	2.97%
maar 'only'	7	1.89%
alleen 'only'	4	1.08%
nooit 'never'	2	0.54%
other licensers	3	0.81%

Appendix 7: Total hours of recording and total amount of utterances in child-directed Dutch (CHILDES)

Database	Hours of recording	Amount of utterances
BolKuiken (Bol and Kuiken 1990)	47	13,966
Groningen (Wijnen and Bol 1983)	170	180,929
VanKampen (Van Kampen 1994)	91.5	58,587
Wijnen (Wijnen 1988, 1992; Wijnen and Elbers 1993)	723	11,751
TOTAL	331.5	265,233

References

Bates, Elizabeth & Brian MacWhinney. 1987. Competition, variation, and language learning. In Brian MacWhinney (ed.), *Mechanisms of language acquisition*, 157–193. Hillsdale, NJ, England: Lawrence Erlbaum Associates.

Baker, C. Carl Lee & John Joseph McCarthy (eds). 1981. *The logical problem of language acquisition*. Cambridge MA: MIT Press.

Berwick, Robert. 1985. *The acquisition of syntactic knowledge*. Cambridge MA: MIT Press.

Berwick, Robert C. & Amy S. Weinberg. 1986. *The grammatical basis of linguistic performance: Language use and acquisition*. Cambridge MA: MIT Press.

Bishop, Dorothy V. M. & Catherine Adams. 1990. A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry* 31. 1027–1050.

Bol, Gerard & Folkert Kuiken. 1990. Grammatical analysis of developmental language disorders: A study of the morphosyntax of children with specific language disorders, with hearing impairment and with Down's syndrome. *Clinical Linguistics & Phonetics* 4(1). 77–86.

Bowerman, Melissa. 1982. Reorganizational processes in lexical and syntactic development. *Language Acquisition: The State of the Art*. 319.

Boyd, Jeremy K. & Adele E. Goldberg. 2011. Learning what NOT to say: The role of statistical preemption and categorization in a-adjective production. *Language* 87(1). 55–83.

Braine, Martin DS. 1971. On two types of models of the internalization of grammars. *The Ontogenesis of Grammar*. 153–186.

Brown, Roger. 1973. *A first language: The early stages*. Cambridge MA: Harvard University Press.

Brown, Roger & Camille Hanlon. 1970. Derivational complexity and order of acquisition in child speech. *Cognition and the Development of Language. New York: Wiley* 8.

Cartwright, Timothy A. & Michael R. Brent. 1997. Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. *Cognition* 63(2). 121–170.

Chierchia, Gennaro. 2004. Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. *Structures and Beyond* 3. 39–103.

Chierchia, Gennaro. 2006. Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic Inquiry* 37(4). 535–590.

Chomsky, Noam. 1972. *Problems of knowledge and freedom: the Russell lectures*. Fontana.

Chomsky, Noam. 1975. Reflections on language. *New York: Pantheon* 212.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, Noam & Howard Lasnik. 1977. Filters and control. *Linguistic Inquiry* 8(3). 425–504.

Clark, Eve V. 1987. The principle of contrast: A constraint on language acquisition. *Mechanisms of Language Acquisition* 1. 33.

Clark, Eve V. 1995. *The lexicon in acquisition*. Cambridge: Cambridge University Press.

Cowie, Fiona. 1997. The logical problem of language acquisition. *Synthese* 111(1). 17–51.

Drenhaus, Heiner, Douglas Saddy & Stefan Frisch. 2005. Processing negative polarity items: When negation comes through the backdoor. *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*. 145–165.

Drenhaus, Heiner, Joanna Blaszczak & Juliane Schütte. 2006a. An ERP study on the strength of licensers in negative polarity constructions. *Under Submission*.

Drenhaus, Heiner, Douglas Saddy & Stefan Frisch. 2006b. Diagnosis and repair of negative polarity constructions in the light of symbolic resonance analysis. *Brain and Language* 96(3). 255–268.

Elbers, Loekie & Frank Wijnen. 1992. Effort, production skill, and language learning. *Phonological Development: Models, research, implications*. 337–368.

Fikkert, Paula. 1994. *On the acquisition of prosodic structure*. Leiden, the Netherlands: University of Leiden dissertation.

Foraker, Stephani & Brian McElree. 2007. The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language* 56(3). 357–383.

Giannakidou, Anastasia. 1997. *The landscape of polarity items*. Groningen, the Netherlands: University of Groningen dissertation.

Giannakidou, Anastasia. 1998. *Polarity sensitivity as (non) veridical dependency*. Amsterdam & Philadelphia: John Benjamins Publishing.

Giannakidou, Anastasia. 2002. Licensing and sensitivity in polarity items: From downward entailment to nonveridicality. *Chicago Linguistic Society (CLS)* 38(1). 29–53.

Gold, E. Mark. 1967. Language identification in the limit. *Information and Control* 10(5). 447–474.

Goldberg, Adele E. 1993. Another look at some learnability paradoxes. Paper presented at the 25th Annual Stanford Child Language Research Forum, Stanford University, 16–18 April.

Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Goldberg, Adele E. 2009. The nature of generalization in language. *Cognitive Linguistics* 20(1). 93–127.

Goldberg, Adele E. 2011. Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics* 22(1). 131–153.

Goldsmith, John. 1976. *Autosegmental phonology*. Bloomington: Indiana University Linguistics Club.

Goldsmith, John. 1980. Meaning and mechanism in grammar. *Harvard Studies in Syntax and Semantics* 3. 423–428.

Grimshaw, Jane. 1981. Form, function and the language acquisition device. In Carl L. Baker & John J. McCarthy (eds.), *The logical problem of language acquisition*, 165–182. Cambridge MA: MIT Press.

Gropen, Jess, Steven Pinker, Michelle Hollander & Richard Goldberg. 1991. Affectedness and direct objects: The role of lexical semantics in the acquisition of verb argument structure. *Cognition* 41(1). 153–195.

Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Clarendon Press.

Hoeksema, Jack. 1997. Corpus study of negative polarity items. *IV–V Jornades De Corpus Linguistics 1996–1997*. Barcelona: Universitat Pompeu Fabre.

Hoeksema, Jack. 1999. *Aantekeningen bij ooit, deel 2: De opkomst van niet-polair ooit* [Notes on *ooit*, part 2: the emergence of non-polar *ooit*]. *TABU: Bulletin Voor Taalwetenschap* 29. 147–172.

Hoeksema, Jack. 2000. Negative polarity items: Triggering, scope and c-command. In Laurence Horn & Yasuhiko Kato (eds.), *Negation and polarity: Syntactic and semantic perspectives*, 115–146, Oxford: Oxford University Press.

Hoeksema, Jack. 2012. On the grammaticalization of negative polarity items. *The Twentieth Annual Meeting of the Berkeley Linguistics Society: General Session Dedicated to the Contributions of Charles J. Fillmore* 20(1). 273–282.

Homer, Vincent. 2012. Domains of polarity items. *Journal of Semantics*. (forthcoming)

Horn, Laurence R. 1989. *A natural history of negation*. Chicago: University of Chicago Press.

Iatridou, Sabine & Hedde Zeijlstra. 2010. On the scopal interaction of negation and deontic modals. *Logic, language and meaning*, 315–324, Springer Berlin Heidelberg.

Iatridou, Sabine & Hedde Zeijlstra. 2013. Negation, polarity and deontic modals. *Linguistic Inquiry* 44(4). 529–568.

Jacobs, Joachim. 1980. Lexical decomposition in Montague Grammar. *Theoretical Linguistics* 7. 121–136.

Kadmon, Nirit & Fred Landman. 1993. Any. *Linguistics and Philosophy* 16(4). 353–422.

Klima, Edward S. 1964. Negation in English. *The Structure of Language*, 245–323.

Koster, Charlotte & Sjoukje Van Der Wal. 1996. Acquisition of negative polarity items. *Child Language. Multilingual Matters, Ltd*. 92–102.

Kratzer, Angelika. 1981. The notional category of modality. In Hans-Jürgen Eikmeyer & Hannes Rieser (eds.), *Words, Worlds, and Contexts: New Approches in Word Semantics*, 38–74. Berlin: Walter de Gruyter.

Kratzer, Angelika. 1991. Modality/conditionals. In Arnim Von Stechow & Dieter Wunderlich (eds.), *Semantik: Ein internationales handbuch der zeitgenossischen forschung* [Semantics: an international handbook of contemporary research], 639–656. Berlin: Walter de Gruyter.

Krifka, Manfred. 1994. The semantics and pragmatics of weak and strong polarity items in assertions. *The Semantics and Linguistic Theory Conference (SALT)* 4. 195–219.

Ladusaw, William. 1979. *Negative polarity items as inherent scope relations*. Austin, the United States: University of Texas at Austin dissertation.

Leonard, Laurence B. 2000. *Children with specific language impairment*. Cambridge MA: MIT press.

Levelt, Clara Cecilia. 1994. *On the acquisition of place*. Leiden and Amsterdam, the Netherlands: Holland Institute of Generative Linguistics dissertation.

Lin, Jing. 2011a. *Mandarin shenme as a negative polarity item: Distribution, interpretation and acquisition*. Amsterdam, the Netherlands: The University of Amsterdam rMA thesis.

Lin, Jing. 2011b. Mandarin *shenme* as a negative polarity item: A semantic approach. Paper presented at the 9th Semantics in the Netherlands Day (SiN IX), University of Utrecht, 11 November.

Lin, Jing & Hedde Zeijlstra. 2012. Mandarin *shenme* as a negative polarity item: Evidence from first language acquisition. *The International Conference of Linguistic Evidence: Empirical, Theoretical and Computational Perspectives* 5. 217–276.

Lin, Jing, Fred Weerman & Hedde Zeijlstra. In prep. A learning path from lexical frames to abstract knowledge: experimental evidence from Dutch children's acquisition of NPI modal *hoeven* 'need'.

MacWhinney, Brian. 2002. The gradual emergence of language. In Talmy Givón & Bertram F. Malle (eds.), *The Evolution of Language Out of Pre-Language*, 231–263. Amsterdam: John Benjamins Publishing.

MacWhinney, Brian. 2009. The childes project. *Tools for Analyzing talk–electronic edition 2*.

Manzini, M. Rita & Kenneth Wexler. 1987. Parameters, binding theory, and learnability. *Linguistic Inquiry* 18(3). 413–444.

Marcotte, Jean-Philippe. 2005. *Causative Alternation Errors in Child Language Acquisition*. Stanford, the United States: Stanford University dissertation.

Marcus, Gary F. 1993. Negative evidence in language acquisition. *Cognition* 46(1). 53–85.

Marcus, Gary F., Steven Pinker, Michael Ullman, Michelle Hollander, T. John Rosen, Fei Xu & Harald Clahsen. 1992. Overregularization in language acquisition. *Monographs of the Society for Research in Child Development* 57(4). i–178.

McCawley, James D. 1988. *The syntactic phenomena of English*. Chicago: University of Chicago Press.

McNeill, David. 1966. Developmental psycholinguistics. In Frank Smith & George A. Miller (eds.), *The Genesis of Language: A Psycholinguistic Approach*, 15–84. Cambridge MA: MIT press.

Mintz, Toben H. 2002. Category induction from distributional cues in an artificial language. *Memory & Cognition* 30(5). 678–686.

Mintz, Toben H. 2003. Frequent frames as a cue for grammatical categories in child directed speech. *Cognition* 90(1). 91–117.

Mintz, Toben H., Elissa L. Newport & Thomas G. Bever. 1995. Distributional regularities of grammatical categories in speech to infants. *The North East Linguistics Society* 25. 43–54.

Mintz, Toben H., Elissa L. Newport & Thomas G. Bever. 2002. The distributional structure of grammatical categories in speech to young children. *Cognitive Science* 26(4). 393–424.

Nam, Seungho. 1994. Another type of negative polarity item. In Makoto Kanazawa & Christopher J. Piñón (eds.), *Dynamic Polarity and Quantification*, 3–15. Center for the Study of Language (CSLI) Lecture Notes. Stanford: CSLI.

Nippold, Marilyn A. 1990. Concomitant speech and language disorders in stuttering children: A critique of the literature. *Journal of Speech and Hearing Disorders* 55(1). 51.

Nippold, Marilyn A. & Ilsa E. Schwarz, I. E. 2002. Do children recover from specific language impairment? *International Journal of Speech-Language Pathology* 4(1). 41–49.

Oostdijk, Nelleke. 2004. *Het Corpus Gesproken Nederlands: Structuur en bruikbaarheid voor taalonderzoek* [The Corpus of Spoken Dutch: Structure and usefulness for linguistic research].

Pablos, Leticia, Elizabeth Shirley, Kepa Erdocia & Douglas Saddy. 2011. Licensing of negative polarity items in Basque: An ERP study. Paper presented at the *10th International Symposium of Psycholinguistics*, Donostia-San Sebastian, 13–16 April.

Penka, Doris. 2011. *Negative indefinites*. Oxford: Oxford University Press.

Penka, Doris & Hedde Zeijlstra. 2005. Negative indefinites in Dutch and German. Paper presented at the 20th Comparative Germanic Syntax Workshop, Tilburg University, 9–11 June.

Penka, Doris. 2012. Split scope of negative indefinites. *Language and Linguistics Compass* 6(8). 517–532.

Pinker, Steven. 1979. Formal models of language learning. *Cognition* 7(3). 217–283.

Pinker, Steven. 1984. *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Pinker, Steven. 1995. Language acquisition. *Language: An Invitation to Cognitive Science* 1. 135–182.

Pinker, Steven. 2013. *Learnability and cognition: The acquisition of argument structure*. Cambridge MA: MIT press.

Postal, Paul M. 2000. The Ohio lectures on squat. New York, the United States: New York University manuscript.

Redington, Martin, Nick Chater & Steven Finch. 1998. Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science* 22(4). 425–469.

Romberg, Alexa R. & Jenny R. Saffran. 2010. Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science* 1(6). 906–914.

Rullmann, Hotze. 1995. *Geen eenheid* [No Unit]. *TABU* 25(2). 194–197.

Saddy, Douglas, Heiner Drenhaus & Stefan Frisch. 2004. Processing polarity items: Contrastive licensing costs. *Brain and Language* 90(1). 495–502.

Saffran, Jenny R. 2002. Constraints on statistical language learning. *Journal of Memory and Language* 47(1). 172–196.

Saffran, Jenny R. 2003. Statistical language learning mechanisms and constraints. *Current Directions in Psychological Science* 12(4). 110–114.

Santelmann, Lynn M. & Peter W. Jusczyk. 1998. Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition* 69(2). 105–134.

Schütte, Juliane. 2006. On the processing of negative polarity constructions: When the licensor makes the difference. *Central European Student Conference in Linguistics* 1. 24–26.

Snyder, William. 2008. Children's grammatical conservatism: Implications for linguistic theory. In Sano Tetsuya, Mika Endo, Miwa Isobe, Koichi Otaki, Koji Sugisaki & Takeru Suzuki (eds.), *An Enterprise in the Cognitive Science of Language: A Festschrift for Yukio Otsu*, 41–51. Tokyo: Hituzi Syobo.

Stromswold, Karin. 1994. What a mute child tells us about language. New Brunswick, the United States: Rutgers University manuscript.

Thiessen, Erik D. 2011. Domain general constraints on statistical learning. *Child Development* 82(2). 462–470.

Tieu, Lyn Shan. 2009. The acquisition of NPI any in English: A case study. Paper presented at the 34th Boston University Conference on Language Development, Boston University, 6–8 November.

Tieu, Lyn Shan. 2011. On the tri-ambiguous status of 'any': The view from child language. *The Semantics and Linguistic Theory Conference (SALT)* 20. 19–37.

Van der Wal, Sjoukje. 1996. *Negative polarity items and negation: Tandem acquisition*. Groningen, the Netherlands: University of Groningen dissertation.

Van der Wouden, Ton. 1997. *Negative Contexts: Collocation, Polarity and Multiple Negation*. London: Routledge.

Van Kampen, N. Jacqueline. 1994. The learnability of the left branch condition. *Linguistics in the Netherlands* 11(1). 83–94.

Von Fintel, Kai. 1999. NPI licensing, strawson entailment, and context dependency. *Journal of Semantics* 16(2). 97–148.

Von Fintel, Kai & Sabine Iatridou. 2003. Epistemic containment. *Linguistic Inquiry* 34(2). 173–198.

Wijnen, Frank. 1988. Spontaneous word fragmentations in children: Evidence for the syllable as a unit in speech production. *Journal of Phonetics* 16. 187–202.

Wijnen, Frank. 1992. Incidental word and sound errors in young speakers. *Journal of Memory and Language* 31(6). 734–755.

Wijnen, Frank & Gerard Bol. 1993. The escape from the optional infinitive stage. *Language and Cognition* 3. 105–118.

Xiang, Ming, Brian Dillon & Colin Phillips. 2009. Illusory licensing effects across dependency types: ERP evidence. *Brain and Language* 108(1). 40–55.

Von fintel, Hedde. 2011. On the syntactically complex status of negative indefinites. *Journal of Comparative Germanic Linguistics* 14(2). 111–138.

Zwarts, Frans. 1981. Negatief polaire uitdrukkingen I [*Negative Polarity Items I*]. *Glot* 4(1). 35–132.

Zwarts, Frans. 1986. *Categoriale Grammatica En Algebraïsche Semantiek: Een Onderzoek Naar Negatie En Polariteit in Het Nederlands* [Categorical grammar and algebraic semantics: a study of negation and polarity in Dutch]. Groningen, the Netherlands: University of Groningen dissertation.

Zwarts, Frans. 1998. Three types of polarity. In Hamm Fritz & Erhard W. Hinrichs (eds.), *Plurality and quantification*, 177–238. Springer Netherlands.