

Starman or Sterrenman: an acquisitional perspective on the social meaning of English in Flanders

Zenner, Eline; Rosseel, Laura; Speelman, Dirk

Published in:
International Journal of Bilingualism

DOI:
[10.1177/1367006920960816](https://doi.org/10.1177/1367006920960816)

Publication date:
2020

Document Version:
Accepted author manuscript

[Link to publication](#)

Citation for published version (APA):
Zenner, E., Rosseel, L., & Speelman, D. (2020). Starman or Sterrenman: an acquisitional perspective on the social meaning of English in Flanders. *International Journal of Bilingualism*, 25(3), 568-591.
<https://doi.org/10.1177/1367006920960816>

Copyright

No part of this publication may be reproduced or transmitted in any form, without the prior written permission of the author(s) or other rights holders to whom publication rights have been transferred, unless permitted by a license attached to the publication (a Creative Commons license or other), or unless exceptions to copyright law apply.

Take down policy

If you believe that this document infringes your copyright or other rights, please contact openaccess@vub.be, with details of the nature of the infringement. We will investigate the claim and if justified, we will take the appropriate steps.

Starman or Sterrenman: an acquisitional perspective on the social meaning of English in Flanders

Abstract

This paper reports on an experiment conducted with 174 Flemish children in three age groups, viz. first, third and fifth year of primary school, aiming to assess the social meaning children associate with English-sourced lexemes in Belgian Dutch. The children were presented with two versions of a cartoon hero developed specifically for this study: *Sterrenman*, who only uses Dutch words, and *Starman*, who uses English alternatives for 17 content words in the cartoon's script. Relying on extensive pretesting, we adapted standard designs and instrumentation from social psychology (the matched guise technique) and language acquisition research to gauge the children's appreciation of the two heroes, their understanding of the English and Dutch vocabulary used in the script, and their level of language awareness. The key findings are: (1) rather than an incremental increase in favorable social meaning for English from the first to the third age group under scrutiny, we see a decline of the prestige of Dutch in the oldest age group; (2) the children's level of language awareness, receptive vocabulary knowledge and age are associated with their overall appreciation of *Sterrenman* and *Starman*; (3) working with children from various age groups inevitably entails methodological risks. Overall, the method and results of our exploratory study on the acquisition of the social meaning of English lexemes by children and their ensuing progressing bilingualism help us contribute to both the recent socio-pragmatic turn in anglicism research and to the upcoming field of developmental sociolinguistics.

Keywords: social meaning; language contact; English; developmental sociolinguistics; acquisition; matched guise; children

1 An acquisitional perspective on the socio-pragmatic turn in Anglicism research

Globalization has left its imprint not only on Europe's economy and culture, but also on its languages. English serves as the default *lingua franca* for international communication in business and education (e.g. Jenkins, Baker & Dewey 2017), and English-sourced lexemes and phrases enrich the domestic languages of Europe. The linguistic variation and change resulting from this contact with English has triggered quite some scholarly attention, particularly as of the second World War (e.g. Krauss 1958 for German). Initially, the majority of this research on English loanwords and phrases in European languages focused on the structural incorporation of English lexical material in the domestic language lexicon (e.g. Carstensen 1980) and on the creation of databases and inventories of English loanwords for lexicographical purposes (e.g. Görlach 2001). The past decades, however, witnessed a clear shift towards a more socio-pragmatic perspective, prioritizing research on the social meaning and pragmatic implication of using English words and phrases versus domestic language alternatives through corpus studies, experiments and perception studies (e.g. Onysko & Winter-Froemel 2011; Zenner & Kristiansen 2013; Andersen 2014; Peterson & Beers Fägersten 2018). The social meaning of English words and phrases is here broadly defined as all social attributes associated with this particular linguistic feature and its users (Walker et al. 2014).

Despite the upsurge of studies on the socio-pragmatic function of English loanwords and phrases in Europe, research on the social meaning awarded to English by young language users, who according to Berns (2007: 43) are “a key player in the process of globalization”, is rare. This paper aims to address this issue, drawing inspiration from insightful studies on the broader theme of children's acquisition of social meaning found in the upcoming field of “developmental sociolinguistics” and in related research on children from bilingual communities outside of Europe.

Developmental sociolinguistics promotes children's acquisition of the social meaning associated with the different linguistic variants, varieties and codes at their disposal in their speech community to “an inherent and active part” of the general acquisition process (De Vogelaer et al. 2017: 20 following Foulkes, Docherty & Watt 2005). Research in this paradigm has so far largely taken the form of corpus studies patterning variation between standard and vernacular forms in children and their caregivers (Smith, Durham & Richards 2013, Holmes-Elliott 2016, Nardy, Chevrot & Barby 2013), though some experimental research can also be found. Notable examples include Kristiansen (2010), who conducted a perception study verifying 150 Spanish children's (aged 6 to 7; 8 to 9; 12 to 13) ability to locate L1 varieties,

and De Vogelaer & Toye (2017), who charted change in the attitudes toward L1 varieties in Belgian Dutch children between 8 and 18 years of age. Both studies rely on Lambert et al.'s (1960) matched guise technique, in which informants are asked to rate the speakers they hear in a set of recorded speech samples on a number of social criteria, typically along the attitudinal dimensions solidarity, status and dynamism (Zahn & Hopper's 1985 speaker evaluation paradigm).

Outside of the European context, this matched guise technique is also found in sociolinguistic research on bilingual children's use and appraisal of the languages (rather than the language varieties) at their disposal. Bokhorst-Heng & Santos Caleon (2009: 238) study 443 10-year old Singapore children's attitudes towards codeswitching between the Singaporean vernaculars and English. Focusing on minority language transmission, Miller (2017) triangulates data by combining the matched guise test, in which separate Spanish and English samples (no code-mixing) were played, with interviews and proficiency tests for 65 Spanish-English young bilinguals in the American Midwest, linking shifts in language preference with shifts in language dominance. In her study, Miller (2017) crucially studies children from six different age groups, which allows her to chart the evolution in the social meaning awarded to English from kindergarten (mean age 5;8) to the fifth grade of elementary school (mean age 11;1). In Europe, the only study to our knowledge that similarly focuses on bilingual children's attitudes towards minority languages is Hoare (2001) which charts attitudes of school-age children in Brittany towards Breton, Breton-accented French and Standard-accented French using a highly varied methodological toolkit including a questionnaire, matched-guise experiment and two types of interviews.

This paper combines the developmental sociolinguistic emphasis on social meaning with Bokhorst-Heng & Santos Caleon's (2009) attention to samples mixing codes and with Miller's (2017) focus on the evolution in social evaluation across age groups, this in the context of the Englishization of Europe and the ensuing progressing bi- and multilingualism. This way, we aim to contribute to (1) the socio-pragmatic turn in Anglicism research by furthering our understanding of the social meaning awarded to the use of English lexemes and phrases in European languages by targeted social groups; (2) developmental sociolinguistics and attitude research in children who have access to different codes in their community by expanding on the triangulated approach by Miller (2017). Two main research questions are addressed:

- RQ1: which evolution do we find in the social meaning awarded to the use of English lexemes and phrases in Belgian Dutch by elementary school children from three age groups?
- RQ2: which parameters help explain the attested evolution?

Before providing a detailed account of the design of our experiment in Section 3, Section 2 provides a brief introduction to the contact situation under scrutiny, with a particular focus on the position of English in the Belgian Dutch elementary school context. The results of our experiment are presented in Section 4, discussed alongside a conclusion in Section 5.

2 English-Dutch contact in acquisition

Dutch is a pluricentric language with two main national varieties: Netherlandic Dutch (spoken in the Netherlands), and Belgian Dutch, spoken in Flanders, the northern part of Belgium. Belgian Dutch is one of the three official national languages of Belgium, alongside French (spoken in Brussels and Wallonia, the southern part of the country) and German, used in a small area in the east of the country. The strong territoriality-driven language legislation and the country's long-lasting "language battles" are reflected in the largely monolingual family language policies (see Zenner & Van De Mieroop 2019), with Dutch-speaking families in the North (Flanders) and French-speaking families in the South (Wallonia)¹. At the same time, contact with English is visible in nearly all domains of life, where it is shown to be linked amongst others with modernity, fun and youth (Gerritsen et al. 2007: 296 on English advertising; De Decker & Vandekerckhove 2012 on English in youth chat channels; Zenner, Speelman & Geeraerts 2015 on English in a Dutch reality TV show;).

Schooling in Flanders forms an exception to this high openness towards English in the sense that French, rather than English, is the second language offered in the Flemish education system, with tuition typically starting in the fifth year of primary school, for children aged ten. English is the third language taught in school, typically offered only as of the second year of secondary school, for children aged thirteen. Studies reveal that nonetheless Flemish 11-year olds have implicitly acquired a substantial amount of English vocabulary prior to the start of

¹ This statement is disregarding the rapidly growing population of multilingual families consisting of expats and immigrant communities in Belgium, which are not included in this study.

English tuition in schools: De Wilde, Brysbaert & Eyckmans (2019) found a median score of 78 correct test items out of 120 for a receptive vocabulary test administered to 867 children in the final year of primary school, and see similar results in Puimège & Peters (2019). Peters (2018) shows that at least in secondary school high levels of English proficiency are linked to children's exposure to English lexemes and phrases in various genres and domains, primarily mass and social media. Overall, although Flemish children cannot be classified as proper bilingual speakers (in the sense that most of them are not able to hold a conversation in English) and are not per se socialized towards using English by their primary caregivers (see Zenner & Van De Mierop 2019), it seems safe to say that they are in contact with the English language prior to formal tuition in the schooling context, which also impacts their L2 acquisition process.

3 Method

3.1 Design

The standing questions in this paper are which social meaning children of various ages attach to English-Dutch mixing (RQ1) and which parameters help explain any evolution noticeable in this social meaning (RQ2). In designing an appropriate method to tackle these research questions, a choice first needs to be made regarding the type of method used to measure social meaning. Given the fact that we want to avoid responses informed by social desirability (Garrett 2010: 57), given the supposedly closer link of indirectly measured attitudes to language variation and change (Kristiansen 2009), and given the successful application of the method in previous research on children (see above, and also Day 1982; Miller 2017), we opted to build on the well-established speaker evaluation paradigm. As explained above, in speaker evaluation experiments, listener-judges evaluate various aspects of the personality of language users in speech samples that typically differ only in the accent or code the speaker employs.

For our study we adhere to this widespread method to assess the social meaning elementary school children between 6 and 11 years old award to Belgian Dutch enriched with English, but (i) introduce some adjustments to adapt the method to our young respondents; (ii) conduct a series of pretests to maximize the validity of our method (Borgers, de Leeuw & Hox 2000: 72); (iii) include a number of posttests to enhance our grasp on the outcome of the matched guise

experiment. In doing so we aim to provide a template for other researchers for studying the social meaning children award to language variants, varieties and codes.

Specifically, we designed a within-subject experiment² that presents children with two versions of an animated cartoon hero developed for this study: *Sterrenman*, who only uses Dutch, and *Starman*, who uses English alternatives for 17 content words in the cartoon's script. After watching the two versions of the hero, children completed an adapted version of the speaker evaluation survey typically used in matched-guise tests (Zahn & Hopper 1985, and see AUTHOR for a pretest of the design). Additionally, children complete three posttests: a vocabulary test including the Dutch and English target words from the script, a brief language recognition test, and a one-question open answer survey. A within-subject design was selected to allow for a comparison of the answer patterns for *Starman* and *Sterrenman* for individual respondents. Below, we describe each of the instruments in more detail.

3.2 Instrumentation

3.2.1 Video and characters

A traditional matched guise test presents respondents with speech samples that are stable in content and speaker but where the speaker's accent or code changes across samples. A first choice concerns the content itself: what will the speaker in the samples discuss? Typically, the content is kept relatively neutral, to reduce the risk that respondents will evaluate the content rather than the linguistic form (Garrett 2010). When working with children it is at the same time important to find samples that stay close to their interests and can sufficiently trigger their attention, especially for the youngest respondents in our study. Choosing a topic for the video that is embedded in the children's social world additionally allowed us to draft a script with English/Dutch word pairs that we anticipated to be part of the children's vocabulary. For these reasons, we chose to create a sample of a new cartoon hero program, complementing the speech sample with video material to make the task more attractive for children (see Borgers, de Leeuw & Hox 2000: 72 on using visual stimuli to make tasks more concrete and interesting, and see Spence, Rollins & Jerger 2002 for an experiment on voice recognition relying on cartoon

² Ethics approved by the ethics committee for SSH at our university. Parents received consent forms prior to the experiment, and children were debriefed at the end of the study.

characters for their higher familiarity for the test subjects). A critical appraisal of the selected genre follows in Section 5.



Figure 1 – cartoon characters

We created cartoon characters in collaboration with our university's Institute for Media and Learning (LIMEL) and a focus group of six elementary school children, who gave feedback on the interim designs. To avoid interference of the cartoon characteristics with respondents' judgements as much as possible, the color scheme of the video was kept relatively neutral by using greyscales. Additionally, the characters all wear masks to avoid influence of ethnicity. This also allows for replication of the study presented in this paper in various applied contexts (see Section 3.2.2) or in other contact situations. Figure 1 shows the eventual characters in the clip with their names in the Dutch-only guise. The protagonists and his helpers are seen on the left-hand pane, the antagonist and his helpers are located on the right-hand pane. We work with one set of male cartoon heroes to restrict the number of clips offered to the children to two (one Dutch, one English), in an attempt not to overburden the youngest participants. This is also why no filler items were included: Miller (2017) reports that a matched guise experiment with four different sound samples is too much for the attention span of the youngest children. Three arguments then motivate the choice for a male rather than a female superhero: (1) a within-subject combination of male and female voices might plausibly have gender overshadow the linguistic differences between the guises; (2) we assume the prototypical superhero to be male for most respondents; (3) speaker evaluation paradigm experiments have predominantly selected male rather than female voices, which makes a male speaker allow for better comparisons across studies (Grondelaers, van Hout & van Gent 2019). Although we do not

attempt such comparisons here, we follow the prototypical design choice of opting for a male voice.

Based on the script (discussed in Section 3.2.2) an animated clip of the heroes was created that was overlaid with two audio samples, one for each version of the script. Each clip lasts 37 seconds.³ The two audio samples were recorded in a broadcasting studio by a professionally trained radio reporter. The voice quality of the radio reporter was pretested by conducting qualitative interviews with 30 respondents (5 boys and 5 girls per age group) who were presented with random speech samples of the radio reporter (viz. not related to the Sterrenman script). The only recurrent comment was his speech rate, which was felt to be too high. When recording the samples for the current study, special attention was hence paid to speech rate.

3.2.2 *Script with lexical alternations*

Following Bokhorst-Heng & Santos Caleon (2009), we created a monolingual Dutch and a Dutch-English speech sample (see also Grondelaers & Speelman 2015 for an experiment including a Dutch-English guise for an adult audience). The monolingual version of the script only contains Dutch lexemes (see (1)). The items in bold font in (1) were replaced by the English content words listed between square brackets in the mixed guise. The script has been drafted in such a way that possible follow-up studies can include alternations between Dutch discourse markers (in italics) and English alternatives. To assess the likelihood that the English words are recognizably English/foreign, we used Onysko's (2007) grapheme-phoneme correspondence rule (see also Zenner, Geeraerts & Speelman 2012), which is confirmed by a pretest on an adult test sample of 10 respondents, and further corroborated through a manipulation check in the actual experiment (Section 3.2.6).

(1) Dutch Version “Sterrenman”

Hallo jongens [guys]! Ik ben **Sterrenman [Starman]**. Samen met mijn **hulpje [sidekick] Planeetkind [Planetkid]** strijd ik tegen **Duistermaan [Darkmoon]** en z'n **bende [gang]**. *Verdorie*, daar is ie weer!

³ The two versions of the video are shared in supplement 1, together with two no-audio versions (one with and one without lip-sync for our script), to allow interested scholars to overlay audio samples of their choice on the video footage. All supplements are attached as appendices to this manuscript.

Brein [Brains], onze **uitvinder [inventor]**, steekt onze **rugzak [backpack]** snel vol met **snufjes [gadgets]** zodat we aan de **strijd [battle]** kunnen beginnen. *Klaar? Daar gaan we!*

Joepie! We kunnen de **slechteriken [bad guys]** zonder problemen verslaan. *“Oh hemel! Geweldig, Brein [Brains]! Wat heb je dat weer fantastisch gedaan! Je bent m’n held [hero], weet je!”*. Dat verdient een **feestje [party]**! Jullie komen toch ook? **Sterrenkoekjes [star cookies]** voor iedereen! *We zien jullie daar!*

An overview of the seventeen word pairs included in the script is provided in Supplement 2. Translation dictionaries (Dutch-English, English-Dutch) and descriptive dictionaries were used to verify the semantic interchangeability of the items in Supplement 2, which was cross-verified by means of a pretest in which 27 adult respondents rated a list of 30 Dutch-English word pairs (including 13 fillers) for translation equivalence on a 7-point scale (1 “this is a very bad translation equivalent for the Dutch word” to 7 “this is an excellent translation equivalent for the Dutch word”). The list of items included in the pretest, the mean scores for interchangeability and standard deviations are provided in Supplement 3. Thirteen of the seventeen word pairs have a mean score of 6 or higher. Four of the target word pairs have a mean between 5 and 6. These lower scores can be explained by the lack of information offered to the participants on the thematic (*hulpje-sidekick*, M 5.0, SD 1.7), stylistic (*duister-dark* M 5.3, SD 1.3, with “gloomy” as contender; *kind-kid* M 5.8, SD 1.2, with “child” as contender) or regional (*jongens-guys* M 5.5, SD 1.1, with “boys” as UK contender) context of the story.

3.2.3 Overall appreciation of the two superheroes

To gauge the children’s overall appreciation of the cartoon heroes they watched, we asked the question: “would you watch these superheroes on TV?”. This question was asked for both *Sterrenman* and *Starman*, so that children were not forced to choose between the superheroes: it may well be the case that they do not make a distinction based on the auditory input, and like or dislike the two heroes to the same extent. Using a separate scale for each hero avoids creating an ambiguous neutral point on the scale that conflates the liking or disliking of both heroes with not having an opinion at all.

(2) Zou je naar deze superhelden kijken op TV?

Starman:	1	2	3	4	5
Sterrenman:	1	2	3	4	5

A rather complicated matter concerns which rating scales are most fit to offer to children. Miller (2017) relies on smiley faces, but this method has been criticized as the icons themselves might not be considered neutral options by the children (particularly young children might be inclined to always color the smiley face rather than the sad face based on personal preferences in smileys⁴). In looking for reliable scales for children, we surveyed studies on pain monitoring, an area of research with a particular interest in the quest for valid scales for gauging children's attitudes and wellbeing. The most interesting study for our current purposes is von Baeyer et al. (2009), who present three datasets supporting the use of numeric rating scales (NRS) over visual analogues (VAS) for children of 8 years and older.⁵ This is further supported by a pretest we conducted with the 30 respondents (10 per age group) described in Section 3.2.1. In a one-on-one setting with the research assistant, the children were first asked to complete a VAS and a NRS assessing their preference for dogs or cats as a pet, after which the research assistant conducted a small interview to assess the reliability of the VAS and NRS ratings and the children's confidence in the task. More inconsistent answers and interpretation difficulties were found for VAS. Hence, we decided to work with NRS. Following the studies reported in McKenna & Kear (1990), we work with a five-point scale (see (2)).

Although both heroes are rated separately, it is important to keep in mind that our within-subject design entails that we measure relative attitudes of Sterrenman compared to Starman. Hence, when we refer to attitudes towards Dutch or English further on in the text, this should always be understood as relative attitudes towards either languages in comparison to the other.

3.2.4 *Social meaning questionnaire*

⁴ Hall, Hume & Tazzyman 2016 suggest a solution to this issue, based on five degrees of happiness

⁵ Important to note is that the word 'numeric' here refers to the fact that the respondent is provided with numbers on the scale to rate the hero rather than just a line (as in VAS). This does not mean that we should necessarily treat the essentially ordinal variable as a numeric parameter in statistical analyses. See Section 4 for more information.

Crucial to the speaker evaluation paradigm is that respondents evaluate various aspects of the personality of the speakers they were exposed to. The social meaning questionnaire was administered to the children as a means to arrive at a more nuanced interpretation of their answer to the overall appreciation question (Section 3.2.3). The questionnaire follows the same set-up as the overall appreciation question: children are asked to evaluate both *Sterrenman* and *Starman* for a total of 12 items (see Supplement 4) on a 5-point NRS (see (2)). The items are designed to capture the traditional attitude dimensions of status and solidarity, and also include questions regarding dynamism (see e.g. Impe & Speelman 2007, Grondelaers, Van Hout & Speelman 2011 or Rosseel, Speelman & Geeraerts 2019 for research revealing the relevance of these dimensions regarding variation in the Flemish linguascape). We anticipate that the dynamism dimension, which was already included in Zahn & Hopper's (1985) seminal work, will be particularly relevant in our study, as the use of English insertions in Dutch has been linked with dynamic social attributes such as modernity, youth and globalization (see Section 2). At the same time, we might anticipate children to award higher solidarity scores to the Dutch hero, given the "homeliness" of Dutch (Zenner & Van De Mieroop 2019). To this end, traditional questions from the 'adult' matched guise surveys were altered to fit children's social context (see Miller 2017: 108 pointing out issues with young children's understanding of the adult evaluation scales). The focus group of six children presented in Section 3.2.1 were asked for feedback and their interpretation of the questions in preparation of the actual experiment. To assess whether children are merely straight lining (each time randomly selecting the same point on the scale), two similar attributes with positive and negative valency were included: smart/stupid for prestige, and fun/boring for dynamism (see Supplement 4).

3.2.5 *Vocabulary test*

One parameter that could influence children's evaluation of the super hero videos is their understanding of the vocabulary used in the script. To this end, we developed a vocabulary test assessing the children's knowledge of the 17 word pairs included in the script (listed in Supplement 2). A comparison of different productive and receptive vocabulary tests available for (bilingual) children, reported in AUTHOR (2017), revealed that a picture-naming task along the lines of the widely used PPVT III (Peabody Picture Vocabulary Test, Dunn & Dunn 2005) is most fit for the age groups under scrutiny.



Figure 2 – example from the picture naming vocabulary test

However, the standard PPVT III does not include our target items. Hence, we created picture naming test items for each of the 17 Dutch items and each of the 17 English items, complemented with 4 filler items. The test items were offered to the children in random order, with the clear restriction that the English item (e.g. *backpack*) was always tested prior to the Dutch item (e.g. *rugzak*). Children were instructed to select the picture that was the best representative of the object designated by the label. An example of a test item is provided in Figure 2. Scoring was binary: an answer was either correct or incorrect. Supplement 5 provides the entire vocabulary test, and Supplement 6 provides more information on the sanity checks included to assess the feasibility and validity of the test.

The sanity checks were relatively limited as concerns the amount of respondents consulted, but should sufficiently safeguard the reliability of the results. At the same time, a notable complexity concerns the difference in the cognitive abilities of our three age groups: devising a test that is sufficiently challenging for the oldest age group (avoiding boredom) and sufficiently feasible for the youngest age group (avoiding frustration) is not straightforward. As is discussed in Supplement 6, we prioritized the youngest group aiming for a feasible vocabulary test overall, though not without including a handful of more challenging test items to meet the needs of the older children.

3.2.6 Language awareness: recognition of foreign items

A guiding assumption in this experiment is that children might evaluate Starman differently from Sterrenman based on the fact that Starman uses English words whereas Sterrenman only uses Dutch words. Although children might unconsciously favor the English hero, we do include a manipulation check to assess children's overt language awareness. We perform this check in two ways: through a self-developed foreign item recognition test (described here) and through an open answer field (Section 3.2.7).

REEKS 'series'	WOORD 1 'word 1'	WOORD 2 'word 2'	WOORD 3 'word 3'	WOORD 4 'word 4'
1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Table 1 – format of the foreign item recognition test

The recognition test consists of eleven word series of four items. The four items are either (i) four Dutch items (two series); (ii) one English item from our script alongside three Dutch items (three series); (iii) two English items from our script alongside two Dutch items (three series); (iv) one English item from our script alongside one French or Spanish item and two Dutch items (one series); (v) one Spanish or French item alongside three Dutch items (two series). The series are read out loud by the research assistant, and for each series the children are asked to tick a box in a grid (see Table 1 for an example) when they hear an English item. When pretesting the method in the group of 30 children presented in Section 3.2.2, it appeared that some children were convinced they had to tick one foreign word per series, which led to guessing. This was remedied in the actual experiment by explaining explicitly that a series does not need to contain one (or just one) English item. Supplement 7 provides a full inventory of the 11 series presented to the children. Scoring is binary for each of the 10 English words included in the test (recognized as English or not), leading to a maximal test score of 10.

3.2.7 *Language awareness: open answer field*

We additionally gauged children's language awareness by offering an open answer field in which they are asked to motivate their overall appreciation of the two videos they saw. Of crucial interest is whether children make any particular reference to language use or code choice in this open answer field.

3.2.8 *Caregiver questionnaire*

A final parameter that is possibly related to children's evaluation of the two superheroes is their general exposure to English in day-to-day life (see Peters 2018 for the link between exposure and proficiency for secondary school children). To this end, we drafted an extensive questionnaire that was distributed via the schools to the children's primary caregivers, both on paper and via a url. Unfortunately, the response rate was very low: 157 of the 204 questionnaires that were sent out were not completed. Hence, we do not further report on the survey.

3.3 *Procedure*

The same 23-year old female research assistant local to the region of the schools conducted the experiment in all groups, following a predefined script. She introduced the videos as samples of a new cartoon for the national broadcasting company's children's channel, and said the children needed to help her out in selecting the best version of the cartoon. She insisted that this was not a graded test and that there were no right or wrong answers. In each age group, children first watched the two videos, then answered the social meaning questionnaire for the two heroes, followed by the overall appreciation question for the two heroes, the open answer field, the vocabulary test and finally, the language recognition test. The order of presentation of the *Sterrenman* and *Starman* video alternated between groups, but no significant impact of sequence was found in the analyses. Note that children saw both versions of the videos, before moving onto the various tasks.

The third and fifth year children were tested in their classroom and were asked to sit apart to avoid copying of other pupils' responses. They received pencil-to-paper tests for all

components. To ensure maximal understanding of the test by the youngest participants, some alterations in the procedure were made in the first year. The first year children were tested in smaller groups of 6 to 7 pupils by the research assistant in a separate classroom (cf. Miller 2017), following the guidelines in Borgers et al. (2000). For practical reasons, it was impossible to conduct the test in smaller groups for the older children. After completing the experiment, all children were rewarded with a sticker of the superhero.

The procedure was designed to ensure maximal understanding of the test questions in each age group and hence of the reliability of the answers of the young respondents, but it is of course also subject to some drawbacks. First, conducting the experiment in a school context reduces the ecological validity of the results, as this is not the default context where children come in contact with the superhero genre. Second, administering the test in different ways in the first year versus the third and fifth year can potentially affect the reliability of the comparison of the groups. This will be taken into account in the Section 5, where we discuss the results of our study.

3.4 Sample

The experiment was conducted in three schools in the vicinity of the same mid-size town in the central province of Vlaams-Brabant between 20 April and 4 May 2018. In each school, we ran the experiment in the first, third and fifth year of elementary school, viz. with boys and girls of 6-7 (first year), 8-9 (third year), 10-11 (fifth year) years old, leading to an initial total of 204 respondents. Given the low response rate for the survey sent to the parents, we also conducted an interview with the teacher to assess the children's language background, excluding all children who did not exclusively use Dutch as their home language ($n = 23$). This was done to ensure the homogeneity of the sample. Additionally, a number of children who clearly "cheated" on the test (e.g. copying the answers of their neighbor on the vocabulary test) were excluded from the analysis ($n = 7$). The sample used for the analyses as such consists of 174 children. Table 2 reveals a sufficiently balanced distribution over age groups and gender.

	F (female respondents)	M (male respondents)	Total
Y1 (year 1, age 6-7)	25	31	56
Y3 (year 3, age 8-9)	31	29	60
Y5 (year 5, age 10-11)	32	26	58
Total	88	86	174

Table 2 – sample

4 Results

The results of our experiment will mainly be presented through scatterplots that visually represent the individual respondent’s replies, further grouped by age (Y1, Y3 and Y5) and gender (F and M). When comparing the two scores (Sterrenman and Starman) for three age groups and two genders for a relatively limited sample, such a visual representation enables a maximally encompassing interpretation⁶. We first present the results for the overall appreciation of the hero (Section 4.1.1), followed by an exploration of the results for the other components of our experiment: the social meaning questionnaire (Section 4.1.2), the vocabulary test scores (Section 4.1.3), and the assessment of language awareness via the recognition test (Section 4.1.4) and the open answer field (Section 4.1.5). Section 4.2 will discuss these components in light of their relationship with the overall appreciation of the hero through multiple correspondence analysis. All analyses are conducted in R. The full dataset can be consulted on request by contacting the authors.

⁶ The nature of both the response and the predictor variables decrease the reliability of inferential statistical tests on our database. The response variable consists of two five-point rating scales (one for Sterrenman and one for Starman), which pose a double issue: (1) a five-point scale (selected as most interpretable for the children) cannot simply be considered as numeric variable; (2) no techniques are available to straightforwardly compare the two rating scales without losing information. Moreover, as will become clear in the remainder of Section 4, the predictor variables show strong ceiling effects by Y5, posing further complications for the application of inferential statistics. Hence, visual representations of the data are the most reliable approach to analyzing this dataset and will hence be relied on for the analyses. At the request of an anonymous reviewer, these analyses will in some cases be accompanied by footnotes reporting on confirmatory bivariate statistical testing, the value of which should for the reasons stated above and given the multifactorial nature of our data (see Section 4.2) not be overstated.

4.1 Age and gender patterns per variable

4.1.1 Overall appreciation

Figure 3 illustrates the overall appreciation of the two superheroes by our respondents. The six panes represent the answers for the six groups of respondents as shown in Table 2: the three age groups are organized horizontally (Y1 on the left, Y3 in the middle, Y5 on the right), gender is organized vertically (female respondents on top, male respondents below). In each pane, the x-axis represents the 1 to 5 rating for Sterrenman, the y-axis represents the 1 to 5 rating for Starman, for the question “would you watch these superheroes on TV?”⁷. Each dot in the pane represents one of the respondents⁸.

Studying the age effect, a clear evolution is seen from the first to the fifth year. For Y1 respondents (6-7 years old), mainly the outer corners of the graph are used, revealing a lack of nuance in the children’s ratings. Apart from that, anything seems to go: some children really like one hero and not the other (1-5 ratings and 5-1 ratings), they really like both (5-5 ratings) or they really dislike both (1-1 ratings). This pattern is more outspoken for the boys than for the girls, who appear to be a bit more nuanced. This trend for more nuanced replies in the girl group becomes stronger in Y3. The girls’ ratings are scattered throughout the plot: not much consistence across respondents is to be found, yet more nuance is found than in Y1. In the group of Y3 boys, we also see such largely inconsistent replies. Again, more nuance is found than in the group of Y1 boys, but less nuance than in the Y3 girls. Although we are taking a perceptual perspective in this study, this is reminiscent of general sociolinguistic findings in production studies where women take the lead in certain linguistic changes (e.g. Labov 1972; *ibid.* 1990). Perhaps this leading role is mirrored in developing sensitivity for the indexicality of linguistic innovations. In Y5, the agreement between the ratings of boys and girls are highly similar and also reveal more consistency than in the earlier groups: with ratings mainly occupying the top left triangle of the plot, we see respondents who give equal ratings to Starman and Sterrenman (located on the diagonal), respondents who rate Starman higher than Sterrenman (located on

⁷ The visual representation in Figure 3 is the most straightforward way to present and interpret the children’s responses to the two rating scales (see footnote 6). In Section 4.2 we further corroborate this visual inspection of the double rating scale with results for a simplified ternary variable expressing children’s preference for Starman and Sterrenman, derived from the difference scores.

⁸ In Figure 3 and in the other dotplots presented in this paper, a small amount of noise is added to the numeric vectors (“jitter”) by drawing samples from the uniform distribution. This helps better visualize overlapping values for different respondents – in this case e.g. setting apart the six female respondents from Y1 who gave both Starman and Sterrenman rating 5.

the upper side of the diagonal), but hardly any respondents (particularly for the boys) who rate Sterrenman over Starman. For this particular experiment and this particular group of respondents, it hence appears that it is not the attitude towards English that is becoming more favorable over the years, but the attitude towards Dutch becoming less favorable over the years.

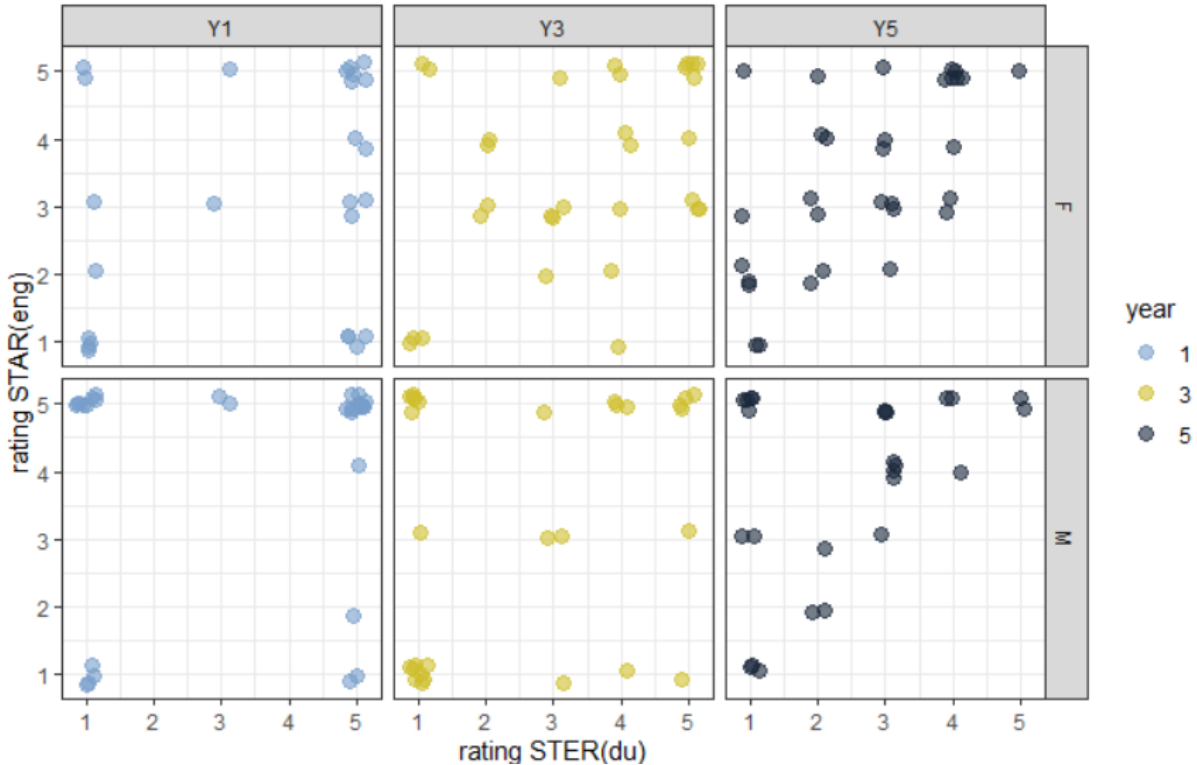


Figure 3 – overall appreciation by age and gender

4.1.2 Components of social meaning

A standing question is then whether we can unravel the social components of the respondents’ overall attitude towards the two heroes. Table 3 provides the standardized Cronbach alpha scores for the ratings of the twelve social attributes provided to our respondents in the social meaning questionnaire, for *Sterrenman* and *Starman* separately. It is clear that the three dimensions of social meaning we envisaged when drafting the scales are not represented in our respondents’ ratings. With each of the 12 alpha scores over .75, it appears that the superheroes are rated on one composite social meaning scale: a positive evaluation of the superhero includes a positive evaluation for all social traits. As no evidence of straight lining was found, we can

assume that the respondents did fill out the questionnaire conscientiously and hence can rule this out as an explanation for not finding the envisaged attitudinal dimensions.⁹

	F		M	
	Sterrenman	Starman	Sterrenman	Starman
Y1	0.94	0.85	0.91	0.87
Y3	0.78	0.88	0.80	0.89
Y5	0.85	0.88	0.88	0.81

Table 3 – standardized alpha for ratings for the questions in the social meaning questionnaire

In light of Figure 3 we are equally (if not more) interested in potential differences in the social rating of Starman versus Sterrenman. Table 4 provides the standardized alpha for the difference scores of the ratings, which was calculated by subtracting the rating for Sterrenman from the rating for Starman. When a respondent rates Starman 5 on the question “would you watch these superheroes on TV?” and Sterrenman 3, this results in a difference score of $5-3=2$.¹⁰ The children’s appreciation is still highly homogeneous, though we witness an incremental decrease in the scales’ consistency over the years: including a fourth age group of boys and girls of 13 years old or above might have revealed some dimensionality in the children’s comparison of the social attributes of the heroes. For now, it seems that the children’s attitudes towards the heroes do not show any notable dimensionality, supporting De Vogelaer & Toye’s (2017: 143) assessment of their youngest informants’ “rather rudimentary attitudes towards language variation” (but see Bokhorst-Heng & Santos Caleon 2009 for more fine-grained dimensions in the attitudes of 10-year-olds in Singapore).¹¹ Note that the alpha scores in Table 4 show the strongest drop in the girls in Y5. Complemented with the findings from De Vogelaer & Toye

⁹ For 11 out of 12 cells in Table 3, the alpha test automatically scores scales for the ratings for the attributes ‘dumb’ and ‘boring’ inversely. The only exception concerns the ratings for *Sterrenman* for the Y3 boys. Here (1) the item whole correlation (corrected for item overlap and scale reliability) for both ‘dumb’ and ‘boring’ are very low (respectively .11 and .12), indicating that these items are not part of the overall scale; (2) ‘dumb’ and ‘boring’ are the only items that when dropped increase the scale’s consistency; (3) only one Y3 boy rates Sterrenman under 3 for both ‘dumb’ and ‘smart’, and only two Y3 boys rate Sterrenman over 3 for both ‘dumb’ and ‘smart’.

¹⁰ Here too, no support for straight lining was found. The ratings for the attribute “dumb” receive an inverse scoring in the Cronbach’s Alpha test in all groups. The ratings for the attribute “boring” receive an inverse scoring in all groups but the Y5 boys. Here, however, “boring” seems to fall outside of the scale for this group: the item whole correlation corrected for item overlap and scale reliability is very low with 0.20, and additionally is the only item in this group that when dropped increases the scale’s consistency.

¹¹ Adding the difference rating for the overall appreciation of the superheroes (Section 4.1.1, Section 4.2) to the scale also does not impact the standardized alpha’s in Table 4 below.

(2017), this is again suggesting they may be somewhat faster in picking up on the social dimensions of language variation.¹²

	F	M
Y1	.94	.94
Y3	.91	.89
Y5	.71	.85

Table 4 - standardized alpha scores for all scales

4.1.3 Vocabulary scores

To gauge our respondents' understanding of the two videos (listening comprehension) a vocabulary test was administered, testing the receptive knowledge of the English and Dutch word pairs included in the script. Figure 4 presents the individual composite test scores for the Dutch (x axis) and English (y axis) vocabulary items (maximum score of 17 for each). The organization of the panes reflects the one in Figure 3¹³.

¹² The steady drop in alpha from Y1 to Y5 encouraged us to attempt factor analysis in all age groups despite the clear warnings from the Cronbach output that not much dimensionality was to be expected. And indeed, no interpretable dimensions were detected through factor analysis.

¹³ Both ANOVA and Kruskal-Wallis indicate that the null hypothesis can be rejected at the .01-level, though see footnote 6.

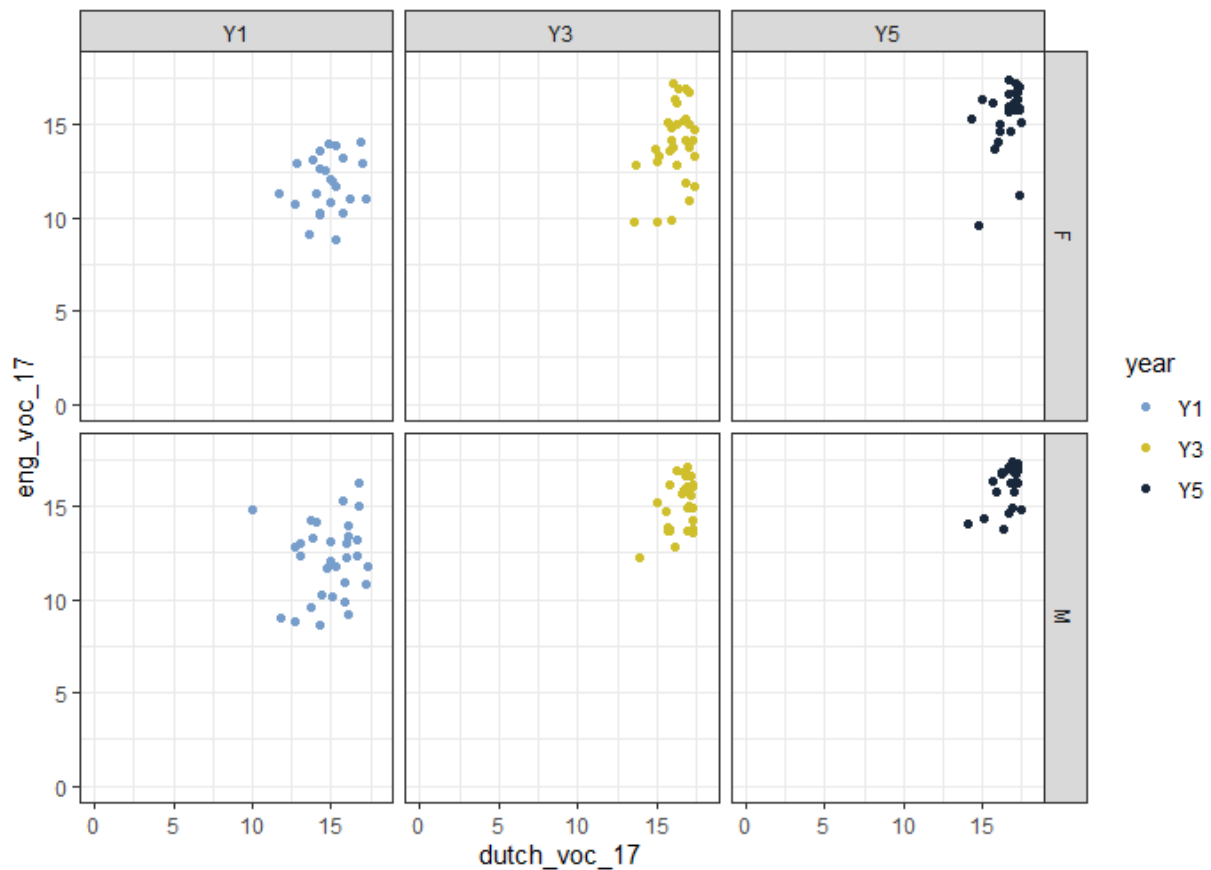


Figure 4 – vocabulary scores by age and gender (maximum test score 17)

Figure 4 reveals a high command of the Dutch vocabulary as of Y1, for both girls and boys ($M = 14.8$, $SD = 1.5$), further increasing across groups (Y3: $M = 16.4$, $SD = 0.84$; Y5: $M = 16.6$, $SD = 0.75$) for both genders, with an outspoken ceiling effect by the fifth year. The decrease in standard deviation over the years reveals an increased homogeneity in the groups. The comprehension of the English words in the script also undergoes a similar evolution across the years, though with more outspoken individual differences in the groups (as reflected in the higher standard deviations) and with a less pronounced ceiling effect in the final group (as reflected in the mean scores, which typically slightly lag behind on the mean scores for Dutch) (Y1: $M = 12.0$, $SD = 1.8$; Y3: $M = 14.6$, $SD = 1.8$; Y5: $M = 15.9$, $SD = 1.4$). Additionally, we see that the difference between the Dutch and English test scores is most outspoken in Y1 and incrementally decreases over the years. The high scores are desirable, of course, as they show that respondents understood the script adding to the validity of the study's outcome. A more extensive error analysis per vocabulary item (presented in Supplement 8) does not reveal any unforeseen issues with the responses and further supports the validity of the test instrument.

4.1.4 Language awareness: recognition

Another factor included in our research design concerns the children's explicit awareness of Starman's use of English. Our first language awareness test gauges children's recognition of 10 of the English vocabulary items used in the script as English. Before proceeding to the results in Figure 5, note that the sample of respondents for this test is slightly smaller than the sample presented in Section 3.4: due to time restrictions the test was not administered to two test groups (N = 13) of the first year pupils. The school was not willing to allow a second testing session, leading to 13 missing values.

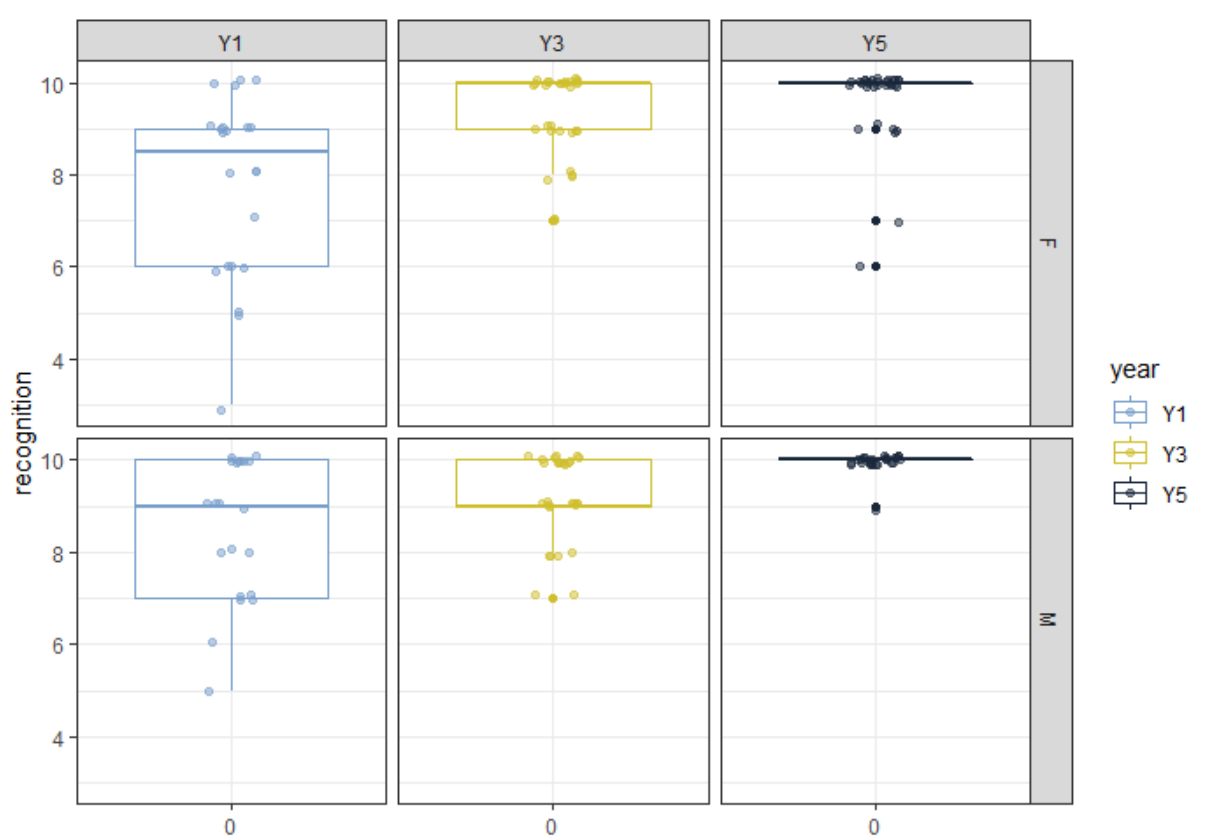


Figure 5 - scores for the recognition test by year and gender (maximum score 10)

The results are reminiscent of the patterns seen for vocabulary: we again witness an incremental change from Y1 to Y5 with increasing test results and decreasing standard deviations, with a

ceiling effect throughout and particularly for Y5, and no notable gender differences (Y1: M = 8.1, SD = 1.8; Y3: M = 9.2, SD = 0.97; Y5: M = 9.8, SD = 0.70).¹⁴

4.1.5 *Language awareness: open answer field*

Turning finally to the open answers provided by the students as a motivation for their overall appreciation of the superhero, qualitative analyses indicate three main themes in the open answers, presented here from revealing more to less awareness of the language used by the superheroes. A first type of answer includes an explicit reference to code choice and English (see (3) and (4))¹⁵. A second type of answer includes a more general reference to language, intelligibility or the lexicon (see (5) and (6)). A third type of answer includes any other type of motivation for the child's overall appreciation (see (7) and (8)). As we will explain in the discussion below, example (8) demonstrates a possible limitation of the superhero genre that we work with: the language variation in the scripts might not be sufficient to modify children's overall appreciation of the superhero genre between guises, which particularly holds for those children with a strong (dis)like of the genre.¹⁶

(3) I'm not very good at English. (id91, F)

(4) Starman is in English and I like that. (id60; M)

(5) Starman has this cool language. (id92; F)

(6) I prefer watching Sterrenman because I understand him better. (id79; M)

(7) I would watch Starman because he is cool. (id113; F)

(8) I don't like watching male superheroes. (id101; F)

Table 5 presents the distribution of the open answers provided by the children over the three answer categories, cross-tabulated with year and gender.

¹⁴ Both ANOVA and Kruskal-Wallis indicate that the null hypothesis can be rejected at the .001-level, though see footnote 6.

¹⁵ We provide the English translation of the Dutch answers in the open comment fields. All examples are drawn from the middle group (Y3). The child's gender and unique ID in the database are provided between brackets.

¹⁶ This also reveals that our current qualitative analyses, which is mainly focused on meta-comments regarding the language use of the two superheroes, might not exhaust the open answer fields. At the same time, given the children's rather short answers, we refrain from further analyses at this point to avoid pushing interpretations onto the data.

type of answer	F						M					
	Y1		Y3		Y5		Y1		Y3		Y5	
	n	%	n	%	n	%	n	%	n	%	n	%
code reference (lang_code)	0	0.0%	6	19.4%	9	28.1%	1	3.2%	6	20.7%	13	50.0%
general comment on language (lang_gen)	6	24.0%	6	19.4%	6	18.8%	7	22.6%	4	13.8%	4	15.4%
no reference to language (nolangref)	19	76.0%	19	61.2%	17	53.1%	23	74.2%	19	65.5%	9	34.6%
Total	25	100%	31	100%	32	100%	31	100%	29	100%	26	100%

Table 5 – distribution of open answers over answer categories

Again a clear evolution is seen across years, with the majority of responses in the first year not containing any reference to language in general or code in particular. An incremental change is seen from the third to the fifth year, which (in contrast to the results for social meaning) seems to happen more outspokenly for the boys than for the girls. As the examples provided in (3)-(8) reveal that each category of answers can support both the choice for *Starman* and the choice of *Sterrenman*, a clear question is how the increased awareness relates to the children's overall evaluation of the two heroes. To this end, the next section turns to a visualization of the relationship between the answer patterns of the respondents for each of the components of this experiment.

4.2 Multiple Correspondence Analysis

Figure 6 presents the output of a multiple correspondence analysis (MCA), a non-inferential technique fit to uncover and visualize underlying structures in a dataset including nominal categorical data (see Chapter 19 in Levshina 2015 for an introduction). Given the double 5-

point rating scales used as response variable and the ceiling effects attested for the receptive vocabulary test and English recognition test, we prefer to work with categorical derivatives from the original categories and feed these into an MCA rather than opt for parametric inferential tests whose reliability would suffer from the inappropriate distributions. In particular, the following seven categorical parameters were included in the MCA: (1) year: Y1, Y3, Y5; (2) gender: M, F; (3) a binary variable derived from the English vocabulary test scores: ‘EV+’ for scores equal to or above the median of 15 vs. ‘EV-’ for scores under the median; (4) a binary variable derived from the Dutch vocabulary test scores: ‘DV+’ for perfect score vs. ‘DV-’ for non-perfect score; (5) a binary variable derived from the English recognition test: ‘EV+’ for perfect score vs. ‘EV-’ for non-perfect scores¹⁷; (6) a ternary variable following the classification of the open answers presented in Table 5; (7) a ternary variable expressing the preference for Starman, derived from the difference score which subtracts the rating for Sterrenman from the rating for Starman, categorized as: STAR (positive ratings, representing higher scores awarded to Starman than to Sterrenman), STER (negative ratings, representing higher scores awarded to Sterrenman than to Starman) and NEUT (neutral ratings, representing equal scores awarded to Starman and Sterrenman) (see Table 6 for a cross-tabulation of this new variable with the factor ‘year’; the table includes the 158 complete responses¹⁸).

	STAR		STER		NEUT	
Y1	11	26.2%	12	28.6%	19	45.2%
Y3	20	33.9%	12	20.3%	27	45.8%
Y5	36	63.2%	3	5.3%	18	31.5%

Table 6 – cross-tabulation of the ternary variable ‘STAR/STER’ and the factor ‘year’

¹⁷ This binary split between perfect and imperfect scores for the Dutch vocabulary and recognition test coincides with a median split for both variables in the full database.

¹⁸ See section 4.1.4 for 13 NA’s for the recognition test. Three other respondents refrained from answering parts of the test. A Chi²-test on Table 6 reveals that the null hypothesis of independence of observations can be rejected at the .001-level, though see footnote 6. The Cramer’s V of 0.248 further reveals a moderate association between ‘year’ and ‘STAR/STER’. When including all respondents, we find a similar pattern (Cramer’s V 0.235, p for Chi² < 0.001).

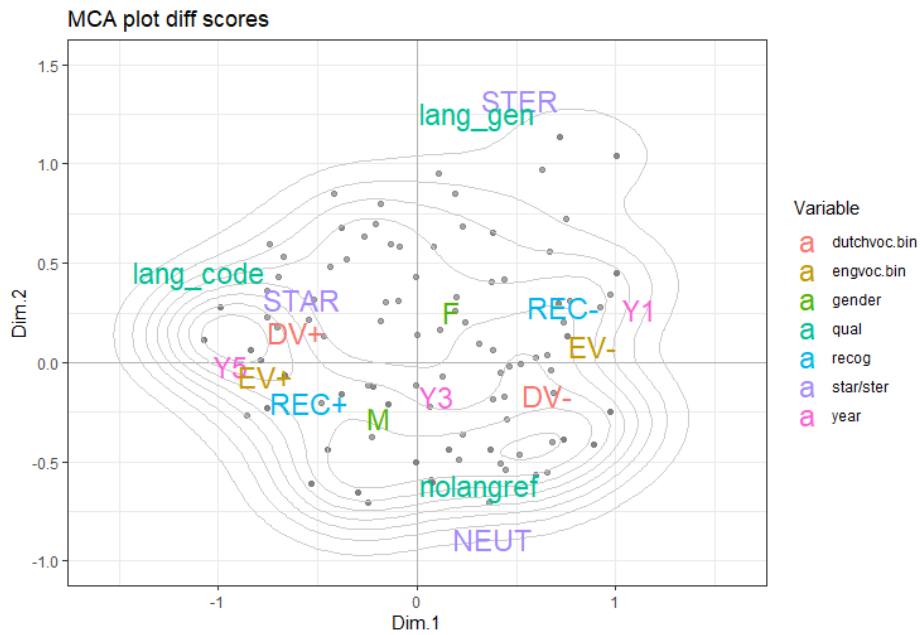


Figure 6 – multiple correspondence analysis (dimension 1: 27.5% variance, dimension 2: 14.3% variance)

Figure 6 reveals the relationship between the seven parameters, with related levels of variables closer together in the two-dimensional space. The dots represent individual respondents¹⁹, with density curves indicating highly concentrated zones in the graph. The two-dimensional solution in Figure 6 covers a very reasonable 41.8% of the variance in the data, and further reveals intuitive relationships between the parameters. The first dimension, which covers 27.5% of the variance, neatly reflects the three age groups involved in the study, with the youngest respondents (Y1) located on the right side of the plot, the middle group (Y3) in the center of the plot and the oldest respondents (Y5) at the left side of the plot. The Y5 group is found alongside STAR (higher ratings for Starman than for Sterrenman), perfect Dutch vocabulary scores (DV+), perfect recognition scores (REC+), English vocabulary scores of 15 out of 17 or higher (EV+) and explicit references to code in the open answer test. Less favorable scores on the vocabulary and recognition tests are in contrast associated clearly with the youngest respondents. The second dimension, which covers 14.3% of the variance, distinguishes between STER (more favorable scores for Sterrenman) and NEUT (equal scores for both superheroes), which are both also located on the right side of the plot and can hence be associated more with Y1 and Y3 than with Y5. Even more revealing is the intuitive link with the open answers scores: the neutral appraisal of Starman and Sterrenman is associated clearly with open answers that

¹⁹ Only complete cases were included.

do not reveal any reflection on the linguistic differences between the guises. Appraisals in favor of *Sterrenman* in contrast are associated with answers concerning language in general. Finally, in line with the results presented in Section 4.1, only minor gender effects are found.

5 Discussion and conclusion

In an attempt to help uncover the evolution in children's acquisition of the social meaning of the language codes and variants available in their speech community, we set up an experiment in which 174 Flemish children from three age groups (first, third and fifth year of primary school) were presented with two versions of a cartoon hero developed specifically for this study: *Sterrenman*, who only uses Dutch, and *Starman*, who uses English alternatives for 17 content words in the cartoon's script. Children were asked to indicate their overall evaluation of the superheroes and rate their social attributes, to comment on this overall evaluation and rating, to complete a vocabulary test and a language recognition test. Our results, presented through graphs prioritizing individual variation in the test results, showed incremental changes from Y1 over Y3 to Y5: the overall evaluation of the superheroes became more nuanced and more favorable for the English hero as children grew older, recognition and understanding of the English vocabulary items increased over the years, with individual differences decreasing, and children's open answers reflecting more awareness of the linguistic differences between *Sterrenman* and *Starman*. At the same time, we find no support for any underlying dimensions structuring our respondents' ratings in the social meaning questionnaire. These findings partially echo results from previous studies that report rudimentary language attitudes in younger children which gradually develop into adult-like structured attitudes as they grow up (cf. Day 1982; De Vogelaer & Toye 2017; Van Bezooijen 1994; Schüppert et al. 2015). The age at which studies report the onset of more refined and structured attitudes seems to differ though, depending on the type of linguistic variation and the speech community. Van Bezooijen (1994) for instance reports significant improvement in the perception of social meaning of regional variation in Dutch for Netherlandic children between the ages of 7 and 10. While the age range covered in our study is similar, attitudes towards English do not seem to develop with the same speed as those for the children in Van Bezooijen's work. It may be the case that attitudes towards L1 variation develop earlier and faster than language attitudes towards other languages, as also suggested by Schüppert et al. (2015) who found that attitudes of Danish and

Swedish children towards the other Scandinavian language start shifting between the ages of 12 and 15. Our findings do not only echo previous work on language attitude development in children, there are also interesting parallels with the development of attitudes towards newly emerging varieties (cf. Kircher & Zipp in preparation). For instance, in their work on Multicultural London English, Kircher and Fox (2019) report one-dimensional attitudes towards this novel variety. Kircher and Fox suggest that new attitudes emerge as monodimensional and only become more complex as people get more experience with the variety in question. We hypothesize that similar cognitive processes could be at play in the development of attitudes towards new varieties in adults as in the development we, and other work before us, have described in children for whom any variety is novel as they are new members of their speech community gradually gaining experience with language variation and acquiring the social norms of their community.

Overall, three key insights can be distilled from our experiment. First, conducting sociolinguistic experiments with children, particularly with children from different age groups, is not straightforward. The differences in cognitive development between our respondents (linked to their social development; see Borgers et al. 2000, De Vogelaer & Katerbow 2017b) make it challenging to design valid instruments and test procedures that accommodate the needs of each age group. Following Borgers et al. (2000), we included several checks and balances to increase the reliability of our results, such as pretesting the materials, including visual stimuli throughout (see Hirschfeld & Gelman 1997) and varying the difficulty of test items to avoid frustration in the youngest group and boredom in the oldest. Yet, some issues unavoidably remain, which need to be addressed in follow-up research. The evaluation of the heroes is for instance subject to genre-effects (see (8)), that need to be accommodated in future work by verifying the appreciation of English characters in less English-prone and/or gender-stereotyped genres. A particularly interesting question for future work in this respect would be to use the case of English words in Dutch to start looking into when and how children develop genre sensitivity in their language attitudes. Furthermore, the attested relationship between children's language awareness and their overall appreciation of the superhero raises questions concerning the precise nature of the experiment we conducted: following the answers in Table 5, our matched guise-inspired experiment can probably still be considered as an indirect attitude measure for the Y1 children (who appear largely unaware of the code difference between Starman and Sterrenman), yet the increased awareness of the Y5 group means that we are to a certain extent triggering evaluations in a more direct fashion in this group. Again, finding a

solution that is at the same time not too cognitively challenging for the Y1 group is not straightforward (cf. the issue of including multiple fillers discussed in Section 3.1.). Implicit attitude measures might however prove a relevant avenue for future research here and can additionally shed light on another layer of the development of children's language attitudes (see Rosseel et al. 2018). More generally, complementing the current findings with studies looking at the same question using different methods may lead to a more profound understanding. In that sense, production data from stylization tasks like role play (cf. Katerbow 2013) may prove an interesting avenue alongside implicit attitude measures to pursue in the future. Next, the results regarding the internal consistency of the difference scores for Starman and Sterrenman indicated that all our respondents operate a homogeneous social scale in evaluating the superheroes, yet also indicated a decrease in the alpha scores over the years. Where an initial reflex might be to include an even older group of respondents (13-14 years old) in the hope of uncovering further developing language attitudes, this again causes methodological issues, given the increased social complexity of this near-adolescent group of respondents (including a school change, see De Vogeleer & Toye 2017: 144). A final limitation of the study that needs to be addressed in the future is the exclusive focus on monolingual children in a low-contact situation. Recent work has shown that attitudes towards English may be very different in a high-contact situation (cf. Miller 2017 on attitudes towards English in English-Spanish bilinguals in the US) and that monolingual and bilingual children may hold different language attitudes towards monolingual and bilingual speakers (cf. Byers-Heinlein and colleagues 2017). Hence, it would be insightful to repeat the current study with a sample of non-monolingual speakers. In short, follow-up research is needed, with different genres, different methods, different age groups, and different contact settings. To inspire researchers interested in such follow-up studies, all materials developed and used for our study are shared in supplements.

The following two key insights relate to our young respondents' understanding of English-Dutch contact as portrayed in this particular genre, which despite the methodological issues described above seem to align with expectations according to the English-Dutch contact situation presented in Section 2. First, studying the children's overall evaluation of the heroes across the years, we found that it is not per se English variants gaining in favorable social meaning: the youngest group already showed a large group of respondents with a favorable attitude towards English (dots in the upper half of the Y1 plots in Figure 3). By contrast, it is the Dutch superhero who is losing prestige by year 5: hardly any children are left who prefer *Sterrenman* over *Starman* (i.e. no dots left in the bottom right corner of the Y5 plots in Figure

3). Second, and finally, the incrementally less favorable attitude towards the Dutch superhero is paralleled by an incremental increase in children's receptive understanding of the English vocabulary presented in the scripts and in their awareness of English as an available code. This finding that perceptual sociolinguistic competence seems to develop in line with language awareness ties in with previous research showing how general language development in young children is closely related to the development of metalinguistic awareness (Smith & Tager-Flusbert 1982). The development of awareness in the context of language attitudes towards English deserves more explicit attention, for instance through open sociolinguistic interviews with children of various age groups: although even the youngest children generally receive high test scores in the English recognition test (see Girard, Floccia & Goslin 2010's results for L2 accent recognition by 5- to 6-year olds), they do not openly comment on Starman's use of English when asked to motivate their overall appreciation of the heroes. So far, however, we have taken some important steps in charting how children grow to "understand the complex social relationships within the communities in which they live". (De Vogelaer & Toye 2017: 118, and see Hirschfeld & Gelman 1997: 214).

6 References

- Andersen, G. (2014). Pragmatic borrowing. *Journal of Pragmatics* 67, 17-33.
- Berns, M. (2007). An empirical approach to the presence of English. In M. Berns, K. de Bot & U. Hasebrink (eds.), *In the Presence of English: Median and European Youth* (pp. 43-51). Boston (MA): Springer.
- Bokhorst-Heng, W. D. & Santos Caleon, I. (2009). The language attitudes of bilingual youth in multilingual Singapore. *Journal of Multilingual and Multicultural Development* 30, 3, 235-251.
- Borgers, N., de Leeuw, E. & Hox, J. (2000). Children as respondents in survey research: cognitive development and response quality. *Bulletin de Méthodologie Sociologique* 66, 60-75.

- Byers-Heinlein, K., Behrend, D. A. Said, L. M. Girgis, H. & Poulin-Dubois, D. (2017). Monolingual and bilingual children's social preferences for monolingual and bilingual speakers. *Developmental Science* 20, 4, e12392.
- Carstensen, B. (1980). Das Genus englischer Fremd- und Lehnwörter im Deutschen. In W. Viereck (ed.), *Studien zum Einfluss der englischen Sprache auf das Deutsche = Studies on the influence of the English language on German* (pp. 37–76). Tübingen: Narr.
- Day, R. (1982). Children's Attitudes Towards Language. In E. B. Ryan & H. Giles (eds.), *Attitudes Towards Language Variation: Social and Applied Contexts* (pp. 116–131). London: Edward Arnold Publishers.
- De Decker, B. & Vandekerckhove, R. (2012). English in Flemish adolescents' computer-mediated discourse: A corpus-based study. *English world-wide* 33(3): 321 – 351.
- De Vogelaer, G. & Katerbow, M. (2017). *Acquiring Sociolinguistic Variation*. Amsterdam: John Benjamins.
- De Vogelaer, G., Chevrot, J.-P., Katerbow, M. & Nardy, A. (2017). Bridging the gap between language acquisition and sociolinguistics. In G. De Vogelaer & M. Katerbow (eds.), *Acquiring Sociolinguistic Variation* (pp. 1-41). Amsterdam: John Benjamins.
- De Vogelaer, G. & Toye, J. (2017). Acquiring attitudes towards varieties of Dutch : a quantitative perspective. In De Vogelaer, Gunther & Matthias Katerbow (eds.) *Acquiring Sociolinguistic Variation* (pp. 117-154). Amsterdam: John Benjamins.
- De Wilde, V., Brysbaert, M. & Eyckmans, J. (2019). Learning English through out-of-school exposure : which levels of language proficiency are attained and which types of input are important? *Bilingualism: Language and Cognition*, in press.
- Dunn, L.M. & Dunn, L.M. (2005). *Peabody Picture Vocabulary Test--Dutch Adaptation (PPVT-III-NL)*(L. Schlichting, Trans.). Amsterdam: Hartcourt Assessment BV.
- Fazio, R. H., & Petty, R. E. (Eds.) (2008). *Attitudes: Their Structure, Function, and Consequences*. New York/Hove: Psychology Press.
- Foulkes, P., Docherty, G. & Watt, D. (2005). Phonological variation in child-directed speech. *Language* 81, 1, 177–206.
- Garrett, P. (2010). *Attitudes to Language*. Cambridge: CUP.

- Gerritsen, M., Nickerson, C., van Hooft, A., van Meurs, F., Nederstigt, U., Starren, M. & Crijns, R. (2007). English in product advertisements in Belgium, France, Germany, the Netherlands and Spain. *World Englishes* 26, 3, 291–315.
- Girard, F., Floccia, C. & Goslin, J. (2010). Perception and awareness of accents in young children. *British Journal of Developmental Psychology* 26, 3, 409-433.
- Görlach, M. (2001). *A Dictionary of European Anglicisms: A Usage Dictionary of Anglicisms in Sixteen European Languages*. Oxford: Oxford University Press.
- Grondelaers, S., Speelman, D. (2015). A quantitative analysis of qualitative free response data. Paradox or new paradigm? In J. Daems, E. Zenner, K. Heylen, D. Speelman & H. Cuyckens (eds.), *Change of Paradigms – New Paradoxes: Recontextualizing Language and Linguistics* (pp. 361-384). Berlin: Mouton de Gruyter.
- Grondelaers, S., Hout, R. van, & Speelman, D. (2011). A perceptual typology of standard language situations in the Low Countries. In N. Coupland & T. Kristiansen (eds.), *Standard Languages and Language Standards in a Changing Europe* (pp. 199–222). Oslo: Novus Press.
- Grondelaers, S., Hout, R. van, & Gent, P. van (2019). Re-evaluating the prestige of regional accents of Netherlandic Standard Dutch. The role of accent strength and speaker gender. *Journal of Language and Social Psychology* 38(2): 215-236.
- Hall, L., Hume, C. & Tazzyman, C. (2016). Five Degrees of Happiness: Effective Smiley Face Likert Scales for Evaluating with Children. In *Proceedings of the The 15th International Conference on Interaction Design and Children* (pp. 311-321). Manchester (UK), June 21 - 24, 2016.
- Hirschfeld, L. A., Gelman, S. A. (1997). What young children think about the relationship between language variation and social difference. *Cognitive Development* 12, 213–238.
- Hoare, R. (2001). An Integrative Approach to Language Attitudes and Identity in Brittany. *Journal of Sociolinguistics* 5, 73–84.
- Holmes-Elliott, S. (2016). Ladies first? Adolescent peaks in a male-led change. *University of Pennsylvania Working Papers in Linguistics* 22, 2, 1-12.

- Impe, L., & Speelman, D. (2007). Vlamingen en hun (tussen)taal: een attitudineel mixed guise-onderzoek. *Handelingen van de Koninklijke Zuid-Nederlandse Maatschappij Voor Taal- En Letterkunde*, 16, 109–128.
- Jenkins, J., Baker, W. & Dewey, M. (eds.). (2017). *The Routledge Handbook of English as a Lingua Franca*. London: Routledge.
- Katerbow, M. (2013). *Spracherwerb und Sprachvariation. Eine phonetisch-phonologische Analyse zum regionalen Erstsprach-erwerb im Moselfränkischen*. Berlin: De Gruyter.
- Kircher, R. and Fox, S. 2019. Attitudes towards Multicultural London English: Implications for attitude theory and language planning. *Journal of Multilingual and Multicultural Development* 40, 10, 847–64.
- Kircher, R. & Zipp, L. (in preparation). An Introduction to Language Attitudes Research. In Zipp, L. & Kircher R. (eds.) *Methods in Language Attitudes Research*. Cambridge: CUP.
- Krauss, P. G. (1958). “The increasing use of English Words in German.” *The German Quarterly* 31, 4, 272–286.
- Kristiansen, G. (2010). Lectal acquisition and linguistic stereotype formation: an empirical study. In D. Geeraerts, G. Kristiansen & Y. Peirsman (eds.) *Advances in Cognitive Sociolinguistics*. Berlin/New York: Mouton de Gruyter.
- Kristiansen, T. (2009). The macro level social meaning of late modern Danish accents. *Acta Linguistica Hafniensia* 41,1, 167–192.
- Labov, W. (1972). The Social Stratification of (R) in New York City Department Stores. *Sociolinguistic Patterns*, 43-69.
- Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, 2, 2, 205-254.
- Lambert, W.E., Hodgson, R.C., Gardner, R.C. & Fillenbaum, S. (1960). Evaluational reactions to spoken language. *Journal of Abnormal and Social Psychology* 60, 1, 44-51.
- Levshina, N. (2015). *How to do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam: Benjamins.
- McKenna, M. C. & Kear, D. J. (1990). Measuring Attitude toward Reading: A New Tool for Teachers. *The Reading Teacher* 43, 9, 626–39.

- Miller, L. (2017). The relationship between language proficiency and language attitudes. Evidence from young Spanish-English bilinguals. *Spanish in Context* 14, 1, 99-123.
- Nardy, A., Chevrot, J.-P. & Barbu, S. (2013). The acquisition of sociolinguistic variation : Looking back and thinking ahead. *Linguistics* 51, 2, 255-284.
- Onysko, A. (2007). *Anglicisms in German: Borrowing, Lexical Productivity and Written Codeswitching*. Berlin: Mouton de Gruyter.
- Onysko, A. & Winter-Froemel, E. (2011). Necessary loans – luxury loans? Exploring the pragmatic dimension of borrowing. *Journal of Pragmatics* 43, 6, 1550–1567.
- Peters E. (2018). The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *ITL - International Journal of Applied Linguistics* 169, 1, 142-168.
- Peterson, E. & Beers Fägersten, K. (2018). Introduction to the special issue: Linguistic and pragmatic outcomes of contact with English. *Journal of Pragmatics*, 133, 105-108.
- Puimège E., Peters E. (2019). Learners' English vocabulary knowledge prior to formal instruction: The role of learner-related and word-related variables. *Language Learning* (epub ahead of print), 1-35.
- Rosseel, L., Speelman, D., & Geeraerts, D. (2019). The relational responding task (RRT): a novel approach to measuring social meaning of language variation. *Linguistics Vanguard*. 5(s1)
- Rosseel, L., Speelman, D. & Geeraerts, D. (2018). Measuring language attitudes using the Personalized Implicit Association Test: A case study on regional varieties of Dutch in Belgium. *Journal of Linguistic Geography*, 6, 1, 20–39.
- Schüppert, A., Hilton, N. H., & Gooskens, C. (2015). Swedish is beautiful, Danish is ugly? Investigating the link between language attitudes and spoken word recognition. *Linguistics* 53, 2, 375-403.
- Smith, J., Durham, M. & Richards, H.M. (2013). The social and linguistic in the acquisition of sociolinguistic variation. *Linguistics* 51, 2, 258-324.
- Smith, C. L. & Tager-Flusberg, H. (1982). Metalinguistic awareness and language development. *Journal of Experimental Child Psychology* 34, 449–468.

- Spence, M. J., Rollins, P.R. & Jerger, S. (2002). Children's recognition of cartoon voices. *Journal of Speech, Language, and Hearing Research* 45, 214–222.
- Van Bezooijen, R. (1994). Aesthetic evaluation of Dutch language varieties. *Language & Communication* 14, 3, 253-263.
- Von Baeyer, C. L., Spagrud, L.J., McCormick, J.C., Choo, E.A., Neville, K. & Connelly, M. (2009) Three New Datasets Supporting Use of the Numerical Rating Scale (NRS-11) for Children's Self-reports of Pain Intensity. *Pain* 143, 3, 223-27. Web.
- Walker, A., García, C., Cortés, Y. & Campbell-Kibler, K. (2014). Comparing social meanings across listener and speaker groups: The indexical field of Spanish /s/. *Language Variation and Change* 26, 02, 169-189.
- Zahn, C. J. & Hopper, R. (1985). Measuring language attitudes: The speech evaluation instrument. *Journal of Language and Social Psychology* 4, 2, 113-123.
- Zenner, E. & G. Kristiansen (eds). 2013. *New Perspectives on Lexical Borrowing: Onomasiological, Methodological and Phraseological Innovations*. Berlin: De Gruyter.
- Zenner E., Speelman D. & Geeraerts D. (2012). Cognitive Sociolinguistics meets loanword research: Measuring variation in the success of anglicisms in Dutch. *Cognitive Linguistics*, 23, 4, 749-792.
- Zenner E., Speelman D. & Geeraerts D. (2015). A sociolinguistic analysis of borrowing in weak contact situations: English loanwords and phrases in expressive utterances in a Dutch reality TV show. *International Journal of Bilingualism* 19, 3, 333-346.
- Zenner E. & Van De Mierop D. (2019). The (near) absence of English in Flemish dinner table conversations. *Applied Linguistics Review* (epub ahead of print), [doi: 10.1515/applirev-2019-0029](https://doi.org/10.1515/applirev-2019-0029).